THE NATURAL HISTORY AND COMPARATIVE TRANSCRIPTOMICS OF A PREDATORY DROSOPHILID, *ACLETOXENUS* CF. *INDICUS*, AND ITS POSITION IN THE DROSOPHILIDAE PHYLOGENY

WONG JINFA

(B.Sc. (Hons), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2015

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Wong Jinfa

Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Rudolf Meier, for giving me the opportunity to further my studies, providing funding for my research and giving me advice throughout the research.

I would also like to thank Mr. Iw Borisuthipandit (Miz), who helped me write the scripts for my phylogenetic analyses. Without his help, I would not even know how to begin the fourth chapter in my thesis. In addition, my thanks to Mr. Diego Pitta de Araujo and Mr. Ng Soon Hwee helped me greatly in the morphology and microscopy work and to Dr. Kathy Su, Dr. Sujatha N. K. and Dr. Amrita Srivathsan for their advice and help.

My deepest appreciation goes to Mr. Foo Maosheng, who were always there to offer his help and to family, friends and guild mates from Serenity Edge who encouraged me whenever I met a road block in my research. I would also like to thank Ms. Gowri Rajaratnam and Mr. Darren Yeo who helped me proof-read my thesis and all the other members of the NUS evolutionary biology lab for their advice and help.

Table of Contents

Summary

Not all species of Drosophilidae have larvae that feed on rotting fruits and plant matter. One such larvae that has be claimed to be a predator of whitefly is from the genus *Acletoxenus*. This was proved to be true where I provide the first video evidence of predation by a Singapore *Acletoxenus* species on *Aleurotrachelus trachoides*. I demonstrated that the larvae have morphological adaptations for predation because the cephaloskeleton lacks the pharyngeal filter and a developed facial mask. Furthermore, I showed that the morphological features that have been traditionally used for distinguishing *Acletoxenus* species (colour patterns on mesonotum) vary across individuals of one population and conclude that the Singapore species is most likely *Acletoxenus* cf. *indicus*. I studied the natural history of *Acletoxenus* cf. *indicus* and find it to have a long development time of 24 days. It was also commonly parasitized by *Pachyneuron leucopiscida* (43% of puparia).

Through performing a comparative analysis of larva transcriptomics, I found that the relatively highly expressed genes in *Acletoxenus* cf. *indicus* are involved in odorant reception, chitin metabolism, alkaloid biosynthesis, and producing glue for pupation. The results also imply that many genes that are only moderately expressed in *Acletoxenus* cf. *indicus* are relatively high expressed in *Drosophila melanogaster* homeostasis and development.

Lastly, I tested whether the phylogenetic relationships of Drosophilidae can be resolved using NCBI GenBank data so that I can place *Acletoxenus* on the tree of life. Two methods (supermatrix and supertree) were used but the results were not satisfactory in terms of resolution, support, and congruence with the existing

classification. The relationships obtained for the Steganinae were very different from previously published phylogenies while there is better congruence for the Drosophilinae. *Acletoxenus* cf. *indicus* was found to be in a "basal" position in the maximum likelihood analyses while it is nested in a clade consisting of Gitonini and *Colocasiomyia* in the maximum parsimony analyses. I conclude that the available data in GenBank are currently insufficient for obtaining a well-resolved and supported phylogenetic hypothesis for Drosophilidae.

List of Tables

List of Figures

Chapter 1. Introduction

The Drosophilidae is a family of flies containing more than 6,500 species over, 131 genera in two subfamilies, the Drosophilinae and Steganinae (Bächli, 2015). The most familiar member in this family would be *Drosophila melanogaster*, which has been used as a model organism in experiments for close to a century (Beckingham *et al.*, 2005). Commonly known as fruit flies, most of its members feed on rotting fruits and plant matter. However, there are exceptions found across different genera that feed on other organisms.

Cladochaeta (Drosophilinae) has ectoparasitic larvae that attach to the abdomens of spittlebug nymphs (Wheeler, 1952; Williams, 1923). The *Drosophila smulivora* species group (Drosophilinae) have aquatic larvae that predate on the eggs and larvae of Simuliidae, Chironomidae and Odonata (Aubertin, 1937; Tsacas & Disney, 1974). *Hirtodrosophila batracida* (Drosophilinae) have larvae that feed on the eggs of the Central American glass frog, *Centrolenella fleischmanni*, although all other members of *Hirtodrosophila* were only known to be mycophagous (Grimaldi, 1994). *Scaptomyza bryani* and other *Scaptomyza titanochaeta* species group (Drosophilinae) larvae found in Hawaii fed on spider eggs in contrast to most other *Scaptomyza* species that are herbivores (O'Connor *et al.*, 2014; Wirth, 1952). *Cacoxenus* larvae (Steganinae: Gitonina) are parasites on bees (Bosch, 1992; Seidelmann, 1999) and *Rhinoleucophenga* (Steganinae) larvae were reported as predators of Sternorrhyncha (Culik & Ventura, 2009). Another drosophilid fly that has been mentioned as a predator of whiteflies was *Acletoxenus* (Steganinae), although there has been no evidence to show that the

interactions were indeed predation and not just associations (Ashburner, 1981; Lambkin & Zalucki, 2010; Malloch, 1929; Parchami-Araghi & Farrokhi, 1995; Yu *et al.*, 2012).

Although Drosophilinae has been extensively studied, there has been very few studies on Steganinae (Otranto *et al.*, 2008). This study, which focused on *Acletoxenus*, would thus add to the lacking knowledge in Steganinae. In Chapter 2, I summarized what we know about *Acletoxenus* based on the current literature. The features used to diagnose each species were reviewed and discussed why they may not be ideal based on the findings of the Singapore population. I also tested the hypothesis that *Acletoxenus* sp. was indeed a predator of whiteflies as previous studies did not provide evidence for this. Finally, the natural history of *Acletoxenus* sp. was explored by studying its morphology, life cycle and behaviour.

As a predator, the genes being expressed by *Acletoxenus* sp. would be very different from that of other Drosophilidae that are mainly saprophagous. In Chapter 3, I looked into the transcriptomes on *Acletoxenus* sp. where I identified and suggested the functions of highly expressed divergent genes in *Acletoxenus* sp. I also compared the expression levels of orthologous genes between the entomophagous *Acletoxenus* sp. and the saprophagous *Drosophila melanogaster*.

Phylogenetic relationships are important as helps us understand the patterns and process of evolution. This allows us to predict if the results of our experiments from one species could be generalized to other species, especially if they are more closely related. Despite numerous laboratories using *Drosophila* in

experiments, the current phylogeny of Drosophilidae remains controversial and poorly studied (Ashburner & Hawley, 2005; Markow & O'Grady, 2006). Attempts to resolve the phylogeny were usually done within subfamilies or with representatives for each genera based on a small number of genes due to computational constraints and lack of data (Otranto *et al.*, 2008; Remsen & O'Grady, 2002; van der Linde *et al.*, 2010; Yassin, 2013). In Chapter 4, I reviewed the current literature on the phylogeny of Drosophilinae. Thereafter, I attempted to resolve the phylogeny of Drosophilidae with the huge dataset from NCBI GenBank as of 7 April 2014, using the latest tree building methods so that I could find out the position of *Acletoxenus* in the Drosophilidae phylogeny. The supertree and supermatrix approaches were carried out based on maximum parsimony and maximum likelihood analyses using 33 gene loci of more than 836 species.

Chapter 2. Identification & Natural History of *Acletoxenus* cf. *indicus*

2.1 Introduction & Literature Review

In 2014, a population of *Acletoxenus* sp. with variable morphology on the mesonotum was discovered in Singapore. This was surprising as the morphology on the mesonotum is unique to each species of this genus according to identification keys (Bock, 1982; Malloch, 1929). This raised the question of whether Singapore's population is composed of different species or is a new morphologically variable species. To answer this question, one would have to review what is currently known about *Acletoxenus*.

The genus *Acletoxenus* was first established by Frauenfeld (1868) based on *Acletoxenus syrphoides.* He used the microtrichose arista as a feature for distinguishing this genus from the other Steganinae genera (Fig. 1). Subsequently, *Acletoxenus syrphoides* Frauenfeld, 1868, was found to have already been described by Leow as *Gitona formosus* in 1864 by Colin (1902). Thus, the name with priority was *Acletoxenus formosus* (Loew, 1864). A few years later, Bock (1982) provided additional features for characterizing the genus (Fig. 1). These include very large eyes, nearly parallel narrow frons, absence of ocellar setae, minute postcellar setae, long orbital setae and absence of carina. The wings have the costa exceeding the apex of R_{2+3} vein, with the cells bm and dm confluent. The legs lack preapical and apical setae (Bock, 1982; Frauenfeld, 1868; Malloch, 1929; Yu *et al.*, 2012).

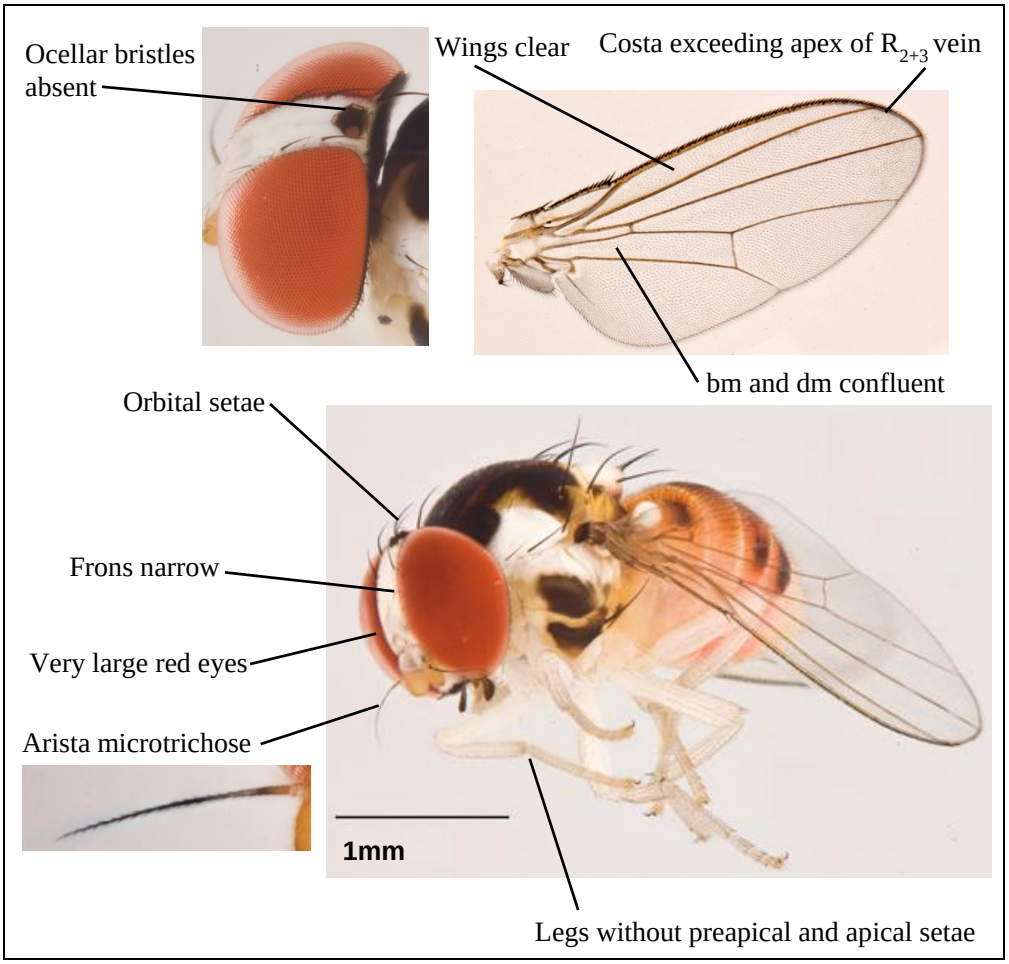

Fig. 1. *Acletoxenus* diagnostic features.

There are currently four valid species of *Acletoxenus* that are described in scientific literature: *Acletoxenus formosus* (Leow, 1864), *Acletoxenus meijerei* Duda, 1924, *Acletoxenus indicus* Malloch, 1929 and *Acletoxenus quadristriatus* Duda, 1936. *Acletoxenus formosus* and *Acletoxenus indicus* were first described based on female specimens although dipterans were usually described based on males (Basden, 1961; Malloch, 1929; Okada, 1977). The holotype for *Acletoxenus meijerei* was lost and the gender is unknown (H de Jong, pers. comm.). The last species, *Acletoxenus quadristriatus*, did not have a holotype assigned and was described based on both sexes (Duda, 1936).

In this genus, the best studied species is *Acletoxenus formosus* (Fig. 2), which has been recorded in southern and central Europe, Israel, Turkey, northern Africa and Australia (Brake & Bächli, 2008). On its head, the proclinate orbital bristles are not noticeably shorter than the anterior reclinate bristles. Malloch (1929) and Bock (1982) both agreed that the mesonotum is largely shining black with the lateral margins yellowish tan. However, Malloch (1929) recorded the mesopleuron to be without a black mark while Bock (1982) and Collin (1902) mentioned a large dark mark at the mesopleuron. The abdomen is pubescent and differ morphologically between those found in Europe and Australia (Bock, 1982). The European males have the first to third abdominal tergites with broadly blackened base while the fourth tergite have a black triangle at the base (Malloch, 1929). The females have markings that were not as extensive and somewhat reduced (Collin, 1902). The Australian individuals are fully tan on the first three tergites. The fourth tergite is tan but black posterolaterally. The fifth tergite is tan with a central black spot and multiple black spots laterally. The sixth tergite has a small central black spot (Bock, 1982).

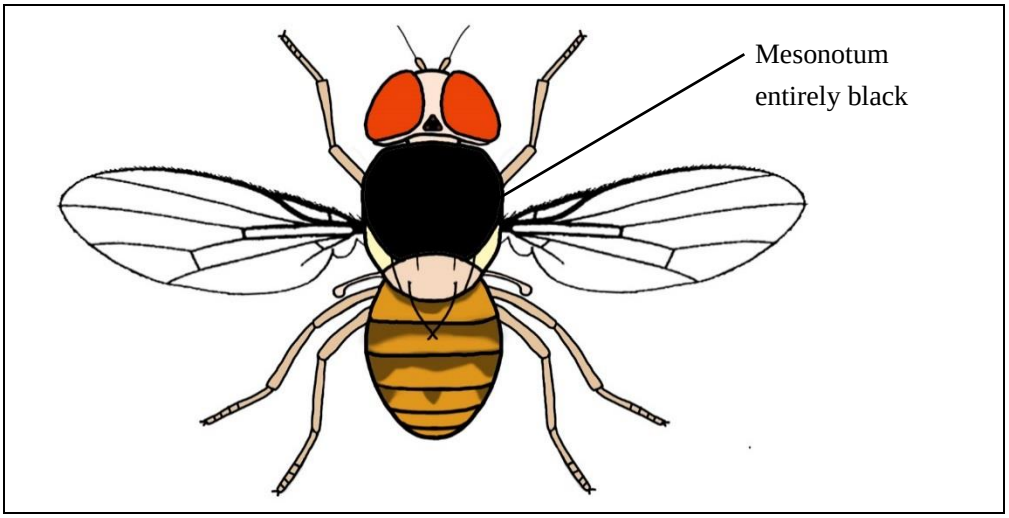

Fig. 2. *Acletoxenus formosus* mesonotum entirely black.

Acletoxenus indicus (Fig. 3) was recorded in eastern to south-eastern Asia (Brake & Bächli, 2008). Malloch (1929) found this species to have proclinate orbital bristles that were very small and fine, not nearly half as long, as the anterior proclinate bristles. The mesonotum has a central black vitta with two other vittas on each side. These two vittas on each side are interrupted at a suture and extended sublaterally. All four vittas do not reach the hind margin. A large black mark is found on the mesopleuron. The abdomen is pubescent with the first to third abdominal tergites with broadly blackened bases that are sometimes reduced to spots. The fourth tergite has a black triangular mark at its base dorsomedially while the fifth had a smaller spot of similar form. Yu *et al*. (2012) measured the body length to be between 1.96–2.22 mm in males and 1.94–2.35 mm in females. They also described the genitalia of the males. The pubescent epandrium with c. 20 short setae protrude ventrally and posterolaterally. The pubescent and setigerous cercus is small, separated from the epnadrium and has three small setae apically on the inner surface. The hypandrium is anteriorly narrow and laterally broad, with five or six setulae per side. It has a narrow anterior rod-like process basally and a lobe-like process bearing four or five setulae submedially. The sclerotized bilobed paramere lacks sensilla. The gonopods are curved dorsally and pointed apicolaterally. The aedagus has two pairs of lobelike processes basally. One is small and the other is curved dorsally that pointed apically where the gracile membranous tube was attached.

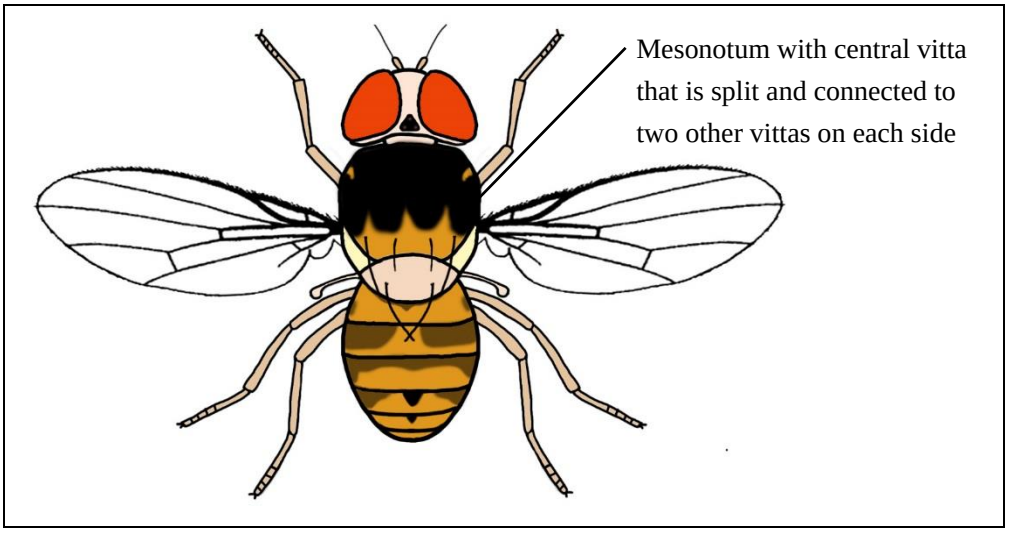

Fig. 3. *Acletoxenus indicus* mesonotum with central vitta that is split and connected to two other vittas on each side

Acletoxenus meijerei (Fig. 4) was described by Duda in 1924, based on material collected in 1909 from Java. Since then, there has been no new sightings of this species. This species has proclinate orbital bristles that are as strong as the anterior proclinate bristles (Duda, 1924; Malloch, 1929). It has a shining yellowish brown thoracic dorsum with two broad vittas which are more or less confluent behind the suture that do not extend to the hind margin (Bock, 1982; Duda, 1924; Malloch, 1929). A dark mark is present on the mesopleuron (Malloch, 1929). The morphology of the abdomen are conflicting in the literature. Malloch (1929) recorded the first to fourth tergites to have a dark spot on each side while Bock (1982) recorded the absence of lateral spots.

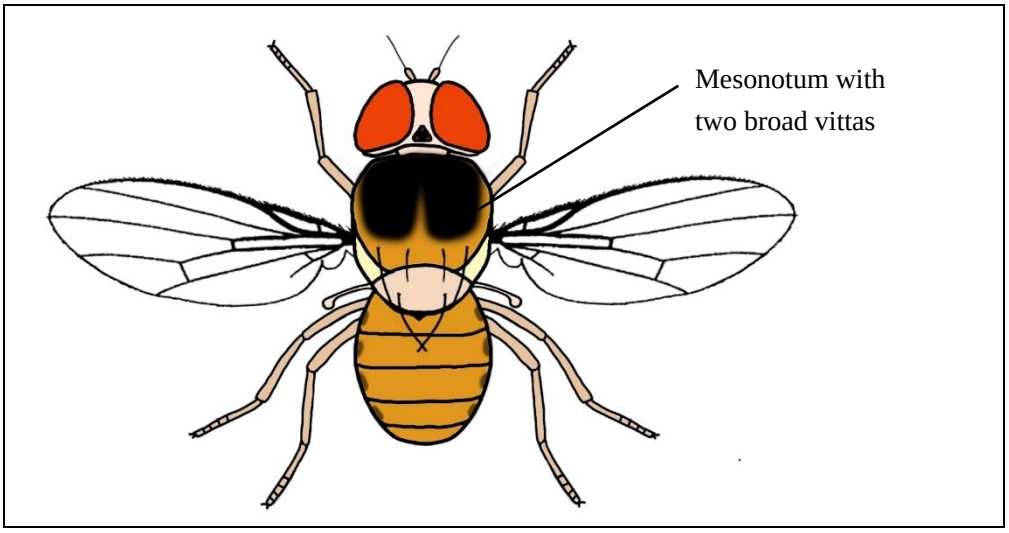

Fig. 4. *Acletoxenus meijerei* mesonotum with two broad vittas

Acletoxenus quadristriatus (Fig. 5) has only been collected from Thursday Island, Torres Strait, Australia (*Acletoxenus quadristriatus* Duda, 1936). The first collection was by Duda in 1934 who described the species and the second was by Dr. Lambkin between 1994 to 2008 during her research on biological control of *Aleurodicus disperses* (Duda, 1936; Lambkin & Zalucki, 2010). Duda (1936) recorded the proclinate orbital bristles to be fine while the reclinate orbital bristles to be strong. In addition, the posterior reclinate orbital bristles are stronger than the anterior reclinate orbital bristles (Bock, 1982). The shining mesonotum has four broad dark brown to black longitudinal vittas coalescing or only slightly separated. The medial vittas reached to rear third of mesonotum while the lateral ones almost to the posterior dorsocentral (Bock, 1982). Duda (1936) made no mention on the morphology of the abdomen except that it was pubescent (Bock, 1982). Dr. Shane McEvey described the syntypes in the Queensland museum to have yellowish abdomen that has small lateral diffused spots on the second, fourth and fifth tergite but no spots on the third tergite. The fifth tergite also has a large dorsomedial round black spot while the sixth tergite

has a smaller spot of similar form. No lectotype has been designated for this species (S McEvey, pers. comm.).

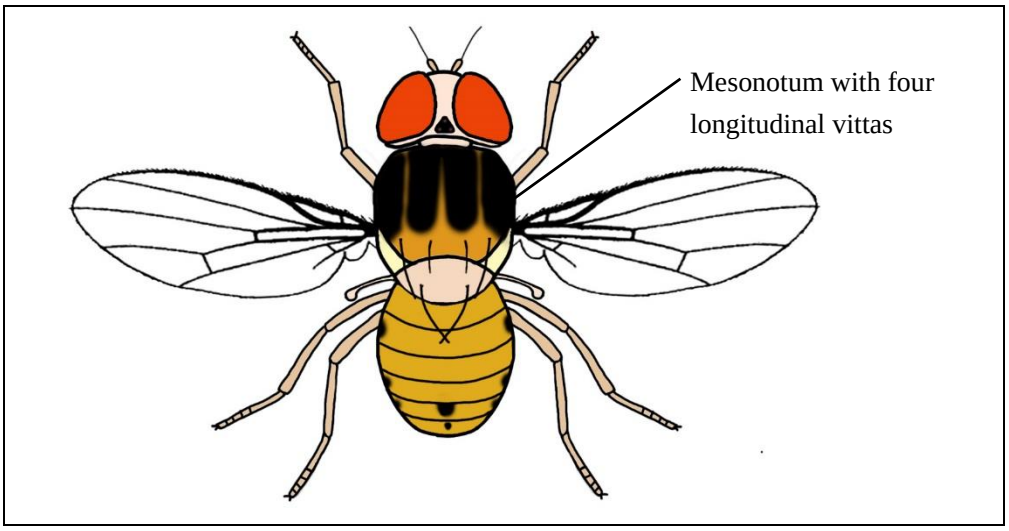

Fig. 5. *Acletoxenus quadristriatus* mesonotum with four longitudinal vittas

There has also been very little studies into the natural history of *Acletoxenus* species, all of which was derived from Clausen and Berry (1932) that is cited by the other papers. Although many papers would mention *Acletoxenus* to be a predator of whiteflies, there has also been no evidence provided to show it was indeed the case and that they were not just associations (Ashburner, 1981; Lambkin & Zalucki, 2010; Malloch, 1929; Parchami-Araghi & Farrokhi, 1995; Yu *et al.*, 2012). Thus, the aim of this study was to:

1. Decipher the species of the Singapore *Acletoxenus*
2. Test the hypothesis that *Acletoxenus* predates on whiteflies
3. Fill up the knowledge gap in Steganinae by studying the natural history of *Acletoxenus*.

2.2 Material and Methods

2.2.1 *Acletoxenus* sp. Recruitment

Chilli (*Capsicum annuum* L. var. *longum* Bailey cv. 'Yang Jiao') seeds were germinated before being transplanted into planter boxes filled with garden soil and Plantaflor® potting mix in a ratio of 3:1. The seedlings were then allowed to grow along a building corridor in the National University of Singapore Kent Ridge Campus at Block S2, Science Drive 4, Republic of Singapore (1° 17' 45.01"N, 103° 46' 41.08"E). Whiteflies were recruited naturally to the chilli plants after two months which in turn attracted *Acletoxenus* sp.

2.2.2 *Acletoxenus* sp. Identification

Adult *Acletoxenus* sp. were captured in the field by hand, and through collection of individuals that were used to determine lifespan (described in Section 2.2.4). After recording gender and morphology of mesonotum by using a dissection microscope, the flies were either stored in 100% ethanol or flash frozen with liquid nitrogen before storing in a freezer at −60ºC. Some samples were sent to Dr. Gerhard Bächli from the Zoological Museum of the University of Zurich and Dr. Shane McEvey from the Australian Museum, who confirmed the fly to be from the genus *Acletoxenus*.

High resolution photographs of the adults were taken with Nikon EOS-1 camera (Visionary Digital) set up. Three morphotypes were identified based on the morphology on the mesonotum corresponding to three species (Bock, 1982; Duda, 1936; Malloch, 1929). The frequency of the three morphotypes were tabulated by sex and Fisher's exact probability 2 × 3 test was used to determine

if the variables were significantly different as less than 80% of the cells had expected frequency that were less than 5. Thereafter, a small-sample test on difference between proportions was used to determine which morphology was significantly different in each sex.

Using the animal tissue protocol, total DNA was extracted from whole specimens of two males and females of each morphotype using QIAGEN® DNeasy® Blood & Tissue Kits. Polymerase chain reaction (PCR) was used to amplify the target mitochondrial cytochrome c oxidase, subunit I (COI) gene using two primer pairs: (1) LCO1490 (5'-GTCAACAAATCATAAAGATAT TGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3'); (2) s2183 (5'-CAACATTTATTTTGATTTTTTGG-3') and a3014 (5'-TCCAAT GCACTAATCTGCCATATTA-3'). The PCR mixture (20 µL) contained 2.5 µL of buffer, 2 µL of dNTP, 1 µL of each primer of a primer pair, 0.15 µL of Ex Taq and 5 µL of template DNA. The program consisted of 40 cycles of amplification (30 sec of denaturation at 94 °C, 30 sec of annealing at 52 °C and 1 min of extension at 72 °C). The PCR products were then purified using BIOLINE SureClean according to the manufacturer's protocol before cycle sequenced using BigDye® Terminator ver. 3.1 Cycle Sequencing Kit. The cycle sequencing mixture (10 µL) contained 2 µL of buffer, 0.5 µL of BigDye®, 1.75 µL of each primer and 2 µL of template DNA. The program consisted of 1 min of initial denaturation at 95 °C, followed by 25 cycles of amplification (30 sec of denaturation at 94 °C, 30 sec of annealing at 52 °C and 4 min of extension at 60 °C). ABI 3730xl sequencer was then used to run out the sequences. Reference COI sequences of *Acletoxenus formosus* (700 base pairs) and *Acletoxenus indicus* (1536 base pairs) were downloaded from GenBank accession numbers

EF576933 and HQ701131 respectively. The sequences of the morphotypes were then aligned against the reference sequences from GenBank using MAFFT ver. 7 using the default settings (Katoh & Standley, 2013). MEGA6 was then used to combine sequences from the same individuals of the morphotypes and used to determine pairwise distances (Tamura *et al.*, 2013).

2.2.3 Are *Acletoxenus* sp. Predators?

Observations of the behaviour of *Acletoxenus* sp. larvae and adults were made in the field and ex-situ. The ex-situ observations consisted of video-taping individuals placed on whitefly infested leaves under a dissection microscope using a Canon LEGRIA HF S30 video camera.

The morphology of the larvae and adults were studied for features that were characteristics of predators. The larvae that were imaged with a scanning electron microscope were killed in hot soapy water before being dehydrated by increasing the ethanol concentrations by 10% every hour until reaching 100%. The adult heads were dissected and then air-dried under a tungsten bulb to fully dehydrate the sample. Thereafter, the dehydrated specimens were coated with platinum before imaging with JEOL JSM 6510 scanning electron microscope.

The cephaloskeleton of the larvae of *Acletoxenus* sp. and *Drosophila melanogaster*, a known saprophage, were dissected for comparison. Upon collection, the larvae were killed in hot soapy water that fully stretched them out for measurement (Alpatov, 1929). A photograph was immediately taken after they were killed with a Nikon EOS-1 camera (Visionary Digital) and used to measure length in Adobe® Photoshop® CS5.5. After photographing, the larvae were dehydrated by increasing the ethanol concentrations by 10% every

hour until reaching 70%. A subset of the larvae were then cut in the mid-section and soaked in potassium hydroxide for 15 minutes or three days. The larvae that were soaked for 15 minutes cleared only the tissue while the larvae that soaked for three days made the cephaloskeleton fully clear for confocal microscopy. The larvae that were soaked for 15 minutes then had their cephaloskeleton removed for imaging under an Olympus BX51 light microscope. The larvae that were soaked for three days were mounted on a glass slide with Euparal before imaging with a Zeiss LSM 510 META confocal microscope at 20 times magnification lens using the 488nm wavelength with a LP505 filter. The images were then rendered into a three dimensional image with Amira 5.3.3. The images from both microscopy methods were then used to compare for differences between the two species of drosophilid flies.

The species of whitefly potentially being predated on by *Acletoxenus* sp. was determined with morphological features on slide mounts as well as COI sequences of fourth instars. Preparation for the slide mounts consisted of first heating up the instars in 80% alcohol for 5 minutes before soaking in 10% potassium hydroxide for three days. The instars were then stained with several drops of glacial acetic acid followed by three drops of acid fuchsin for 20 minutes. These were then washed with 75% ethanol and placed in 95% ethanol for complete dehydration. The instars were then mounted with Euparal and identified with taxonomic keys (Hodges & Evans, 2005; Martin, 1987). Using similar protocols for *Acletoxenus* sp., DNA was extracted and sequenced from the fourth whitefly instars collected off the chilli plant leaves. The only difference in the PCR protocol was the use of a different set of primer pairs, mlCOIintF (5’-GGWACWGGWTGAACWGTWTAYCCYCC-3’) and

jgHCO2198 (5'-TAIA CYTCIGGRTGICCRAARAAYCA-3'). A few of the fourth instars were also sent to a whitefly expert, Dr. Paul De Barro, from the Commonwealth Scientific and Industrial Research Organisation, Australia, for identity confirmation.

2.2.4 Natural History of *Acletoxenus* sp.

Field observations of each stage in the life cycle of *Acletoxenus* sp. was made throughout the experiment. Individuals of each stage were collected and imaged with a Nikon EOS-1 camera (Visionary Digital). The life cycle of *Acletoxenus* sp. was determined by visual inspections on whitefly infected chilli plants out in the field daily because *Acletoxenus* sp. could not be reared ex situ. This also allowed the tracked individuals to develop in a natural setting as much as possible. Upon discovery of an *Acletoxenus* sp. egg, larva or pupa, an individual's length from front to end was measured with a Vernier calliper. The leaf which held the individual was then marked by tying a string with a label to it. On the following day, the leaf that was labelled on the previous day was checked for the same individual to record its stage in the life cycle. If a larva was no longer present on the same leaf as on the previous day, the next closest leaf was checked until a larva was located. The located larva was deemed to be the same from the previous day if its length was the same or longer. If an individual larva could not be located, the data collected for that individual was not used for determining the duration of the larval stage. If there were multiple larvae on the leaf, data was only collected if the length of the larvae were very different so that they could be differentiated. Calculations for each stage were done using the data from an individual's second stage of metamorphosis onwards. For example, if an individual was discovered as a larva, only its

duration as a pupa was used, which minimized underestimation of the duration of each stage. This did not apply to the duration of the egg stage as egg lying events were hard to observe.

To document how adults emerge, a few pupae (n= 34) were collected and placed on moist tissue paper in an enclosed plastic container under a dissection microscope for recording with a Canon LEGRIA HF S30 video camera. The newly emerged adults were then used to determine the life span of *Acletoxenus* sp. adults. Attempts made to rear *Acletoxenus* sp. adults in a container containing whitefly infested chilli plants at room temperature were abandoned as the population of whitefly would increase drastically such that the plants died within three days. Instead, the newly emerged adult flies were reared in Petri dishes. These contained a piece of whitefly-infested leaf placed on a moist piece of tissue paper and a cotton ball soaked in honey (Fig. 6). The leaves were changed every other day while the cotton ball weekly to ensure an adequate supply of food. The lifespan of an individual adult was then determined by counting the number of days from emergence to death. Upon death, the individuals were stored in 100% ethanol for use in identification (Section 2.2.2).

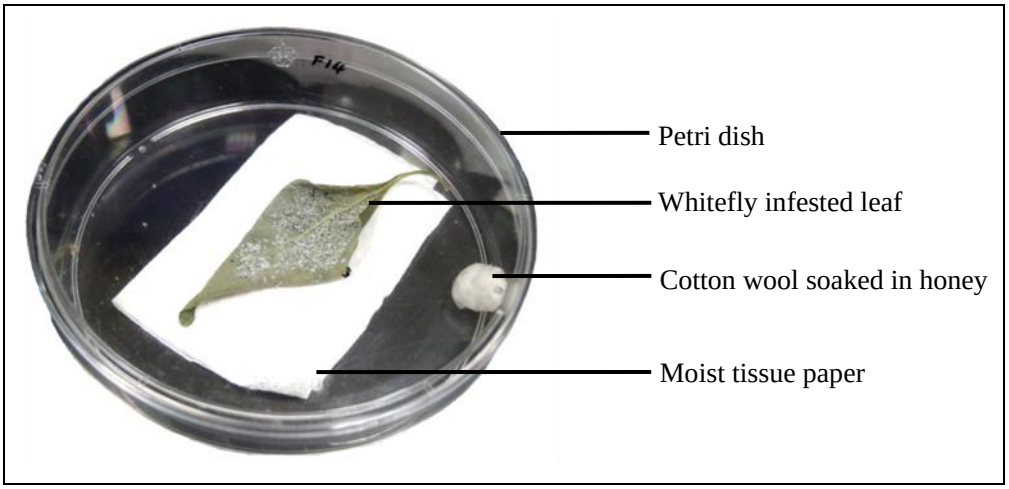

Fig. 6. Petri dish set up used to rear *Acletoxenus* sp. adults

In the last four months of the experiment, the population of *Acletoxenus* sp. declined and many *Acletoxenus* sp. pupae were observed to be black in colour instead of green (Fig. 7). Suspecting the black pupae to be parasitized, a few of them were placed on wet tissue in a plastic container to collect any emerging parasitoids which would then be killed in 100% ethanol and identified with taxonomic keys (Gupta *et al.*, 2013; Noyes, 2015; Sureshan & Narendran, 2000). Photographs of the parasitoid was taken with a Nikon EOS-1 camera (Visionary Digital) and sent to Dr. Hannes Bauron from the Natural History Museum Bern for verification. The parasitoids also had their DNA extracted and the COI sequenced using the same protocol as *Acletoxenus* sp. except that the primer pairs, LepF (5'-ATTCAACCAATCATAAAGATAT TGG-3') and LepR (5'-TAAACTTCTGGATGTCCAAAAAATCA-3'), were used instead during the PCR stage.

Fig. 7. Green *Acletoxenus* sp. pupa (left) and parasitized black pupa (right).

Adult *Acletoxenus* sp. would emerge from the dorsal anterior part of the pupa where a one can easily spot a distinct lid (Fig. 7 & 8). On the other hand, parasitoids would emerge by biting out an opening through any part of the pupa (Fig. 8). Empty pupal cases from the emergence of adult *Acletoxenus* sp. were observed to be translucent but those that had parasitoids emerging had dark brown pigments left in them (Fig. 8). Thus, the whitefly infested chilli plants

were examined for empty pupal cases that were collected in the last three months of study (May, June & July 2014). The monthly rate of parasitism was then estimated by the number of parasitized pupa cases divided by total number of pupa cases collected per month. Thereafter, an average rate of parasitism was calculated for the last three months.

Fig. 8. Empty pupa case of *Acletoxenus* sp. after adult emergence (top row) and pupa case after parasitoid emergence (bottom row).

2.3 Results

2.3.1 Description of Singapore *Acletoxenus* sp.

The Singapore *Acletoxenus* species has proclinate orbital setae that are noticeably shorter than the anterior reclinate setae (Fig. 9). It has variable morphology on the mesonotum and the abdomen. Based on the morphology on the mesonotum, the species in Singapore belonged to three morphotypes: (1) *Acletoxenus quadristriatus* morphotype which has a mesonotum with four dark longitudinal stripes (Fig. 10); (2) *Acletoxenus indicus* morphotype which has a mesonotum with central black vitta that is split and connected to two other vittas on each side (Fig. 11); (3) *Acletoxenus formosus* morphotype which has a mesonotum that was entirely black (Fig. 12).

The *Acletoxenus quadristriatus* morphotype is only present in males while the other two morphotypes are present in both sexes (Fig. 13). Gender and morphotypes were also found to be significantly dependant (Fisher's exact probability test, p-value < 0.01). For males and females, there were no significantly differences between *Acletoxenus quadristriatus* morphotype to *Acletoxenus formosus* morphotype (small-sample test of proportions, p-value = 0.973). There was also no significant difference between the *Acletoxenus quadristriatus* morphotype to *Acletoxenus indicus* morphotype in both genders (small-sample test on proportions, p-value = 0.881). However, females had a significant difference between the *Acletoxenus formosus* morphotype and the *Acletoxenus indicus* morphotype (small-sample test on proportions, p-value = 0.002) that was not significant in males (small-sample test on proportions, p-value = 0.999).

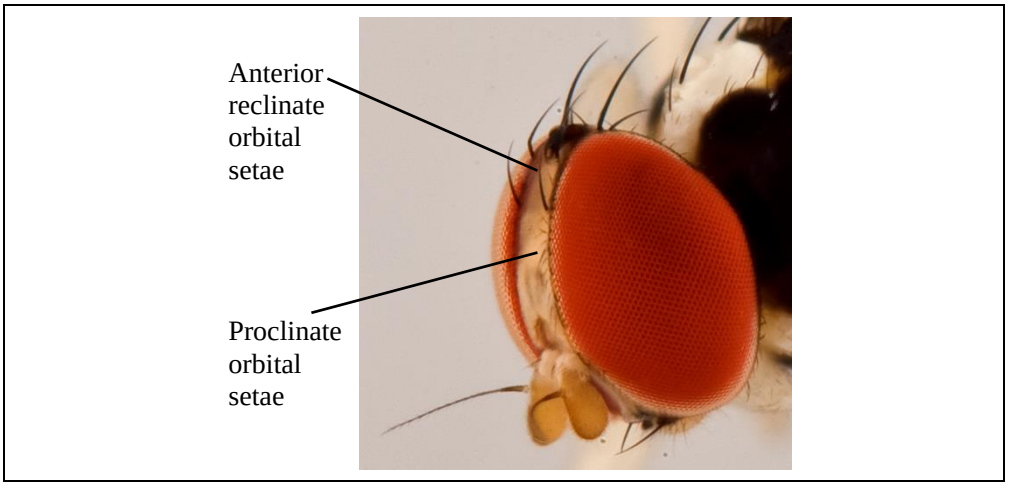

Fig. 9. *Acletoxenus* sp. proclinate orbital setae noticeably shorter than the anterior reclinate setae.

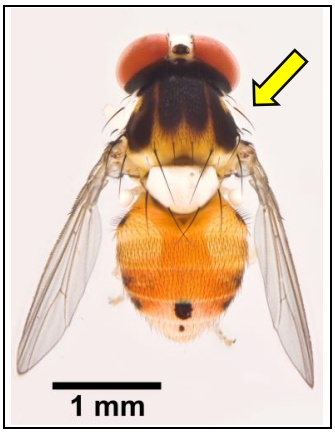

Fig. 10. Mesonotum with four dark longitudinal stripes.

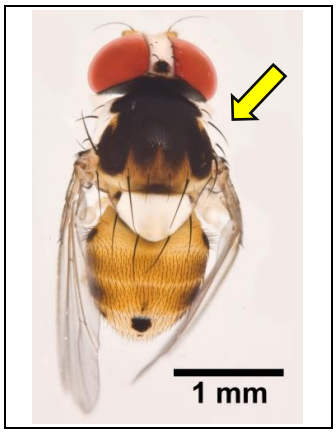

Fig. 11. Mesonotum with central black vitta that is split and connected to two other vittas on each side

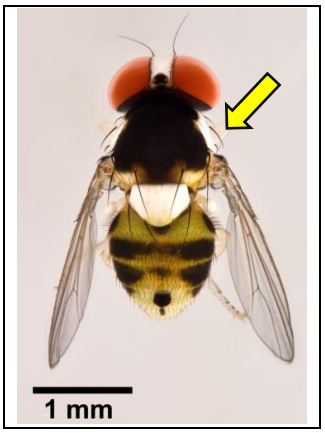

Fig. 12. Mesonotum entirely black.

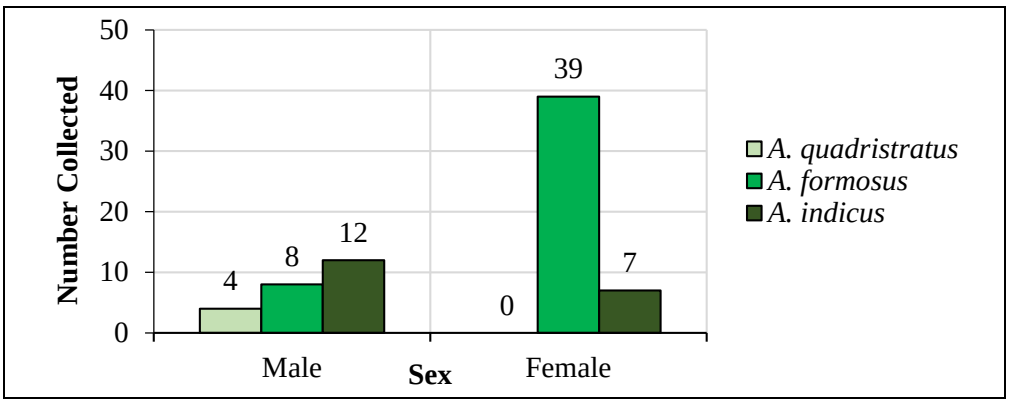

Fig. 13. Number of *Acletoxenus* sp. adults collected based on different morphotypes.

The only morphological feature that is shared between all morphotypes on the abdomen is the presence of a dorsocentral black mark in the fourth tergite and a much smaller mark of similar shape on the fifth tergite (Fig. 10, 11 & 12). The morphology on the other tergites is not consistent within a morphotype or sex and ranged from having broadly blackened base at the base of the tergites (Fig. 12) to reduced spots (Fig 10 & 11). The samples sent overseas were also thought a new species of *Acletoxenus* (G. Bächli, pers. comm.).

The COI region of two individuals of each sex of each morphology are given in Appendix, Table A1 (pg. A-1 to A-3). When these sequences were aligned and compared, the average pairwise distance between the 12 individuals was 0.06%. The pairwise distance was 1.69% when compared to the sequenced of a specimen identified as *Acletoxenus indicus* in Genbank (Accession number: HQ701131.1) and 11.14% when compared to the sequence for *Acletoxenus formosus* (Accession number EF576933.1).

2.3.2 Are *Acletoxenus* sp. Predators?

The first video evidence of the larvae predating on whiteflies was obtained in this study (Available online at https://youtu.be/d8LorGyA_ik). Active during the day, the larvae move through peristaltic contractions of its abdominal segments to approach their stationary whitefly prey. During locomotion, the larvae raise their heads and swing them from one side to the other (as if they are sensing/looking for prey) before placing it down as an anchor while the abdominal segments contract forward. Upon locating prey, the larva uses its mouth hooks to stab a whitefly puparium before sucking up the body fluids of the prey (Fig. 14). Sometimes during feeding, the empty whitefly puparium may

get dislodged from the leaf. These whitefly puparium, in addition to whitefly eggs and wax could become stuck on to the body of the *Acletoxenus* sp. larva by a mucus which it secretes (Fig. 14) (Ashburner, 1981; Clausen & Berry, 1932). Although moving very slowly, the larvae are observed to be capable of moving to different leaves to seek more prey. The larvae also has the ability to move backwards without changing the orientation.

The preoral cavity on the ventral side of the pseudocephalon has few oral ridges flanking the mouth and lack a well-developed facial mask (absence of cirri; Fig.15). The cephaloskeleton of *Acletoxenus* sp. is less sclerotized than that of *Drosophila melanogaster,* as most of the skeleton was already translucent when no potassium hydroxide was used during the dissection process (Fig. 16 & 17). The pharyngeal filter apparatus was not visible in the images of *Acletoxenus* sp. (Fig. 18) but was very visible in those of *Drosophila melanogaster* larvae (Fig. 19).

Fig. 14. *Acletoxenus* sp. larvae feeding on whitefly instar

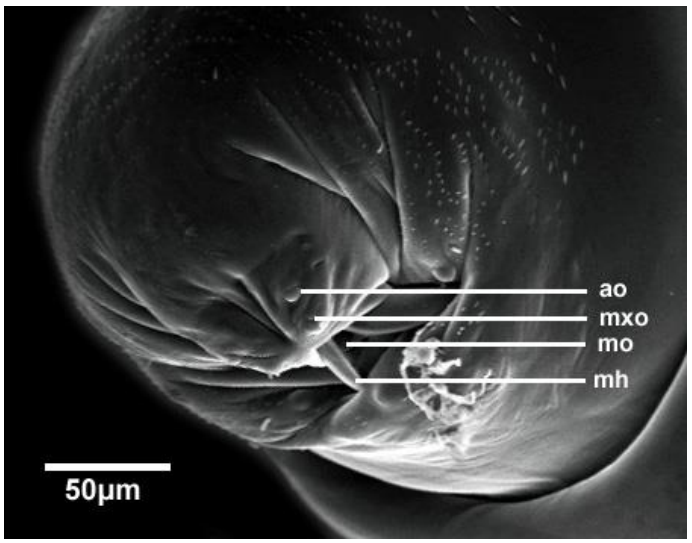

Fig. 15. *Acletoxenus* sp. pseudocephalon. Abbreviations: ao, antennal organ; mxo, maxillary sense organ; mo, mouth opening; mh, mouth hooks.

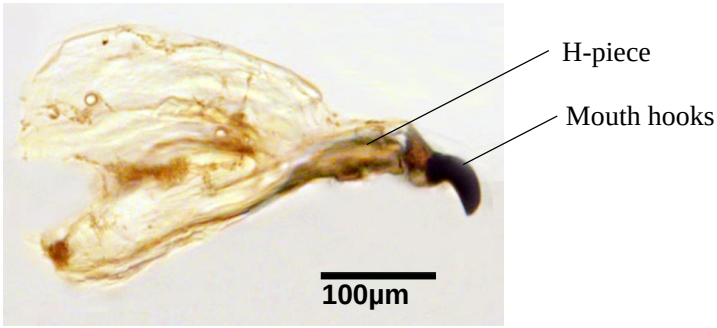

Fig. 16. *Acletoxenus* sp. larva *Acletoxenus* sp. larva cephaloskeleton and mouth hooks (lateral view) without treating with potassium hydroxide.

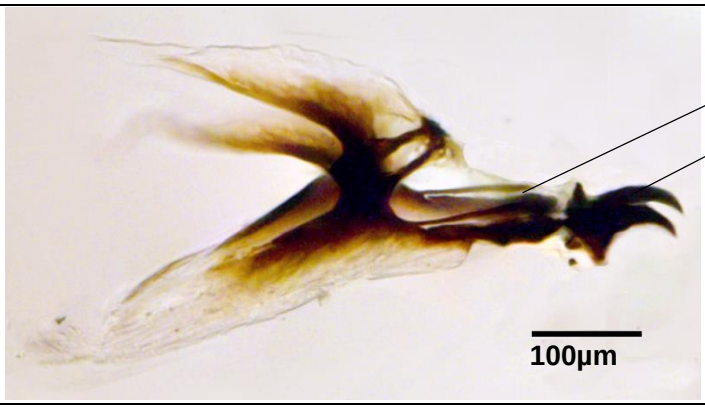

Fig. 17. *Drosophila melanogaster* larva cephaloskeleton and mouth hooks (lateral view) with 30 minutes of potassium hydroxide treatment

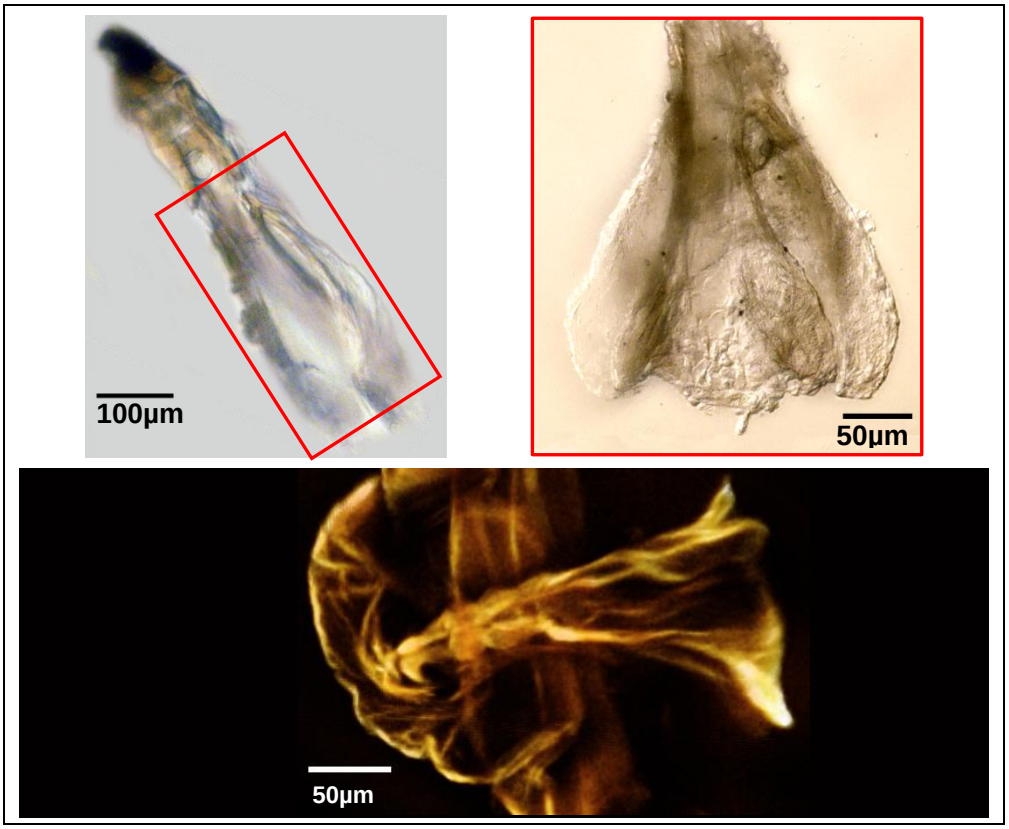

Fig. 18. *Acletoxenus* sp. larva cephaloskeleton ventral view with light microscope (top) and confocal microscope (bottom)

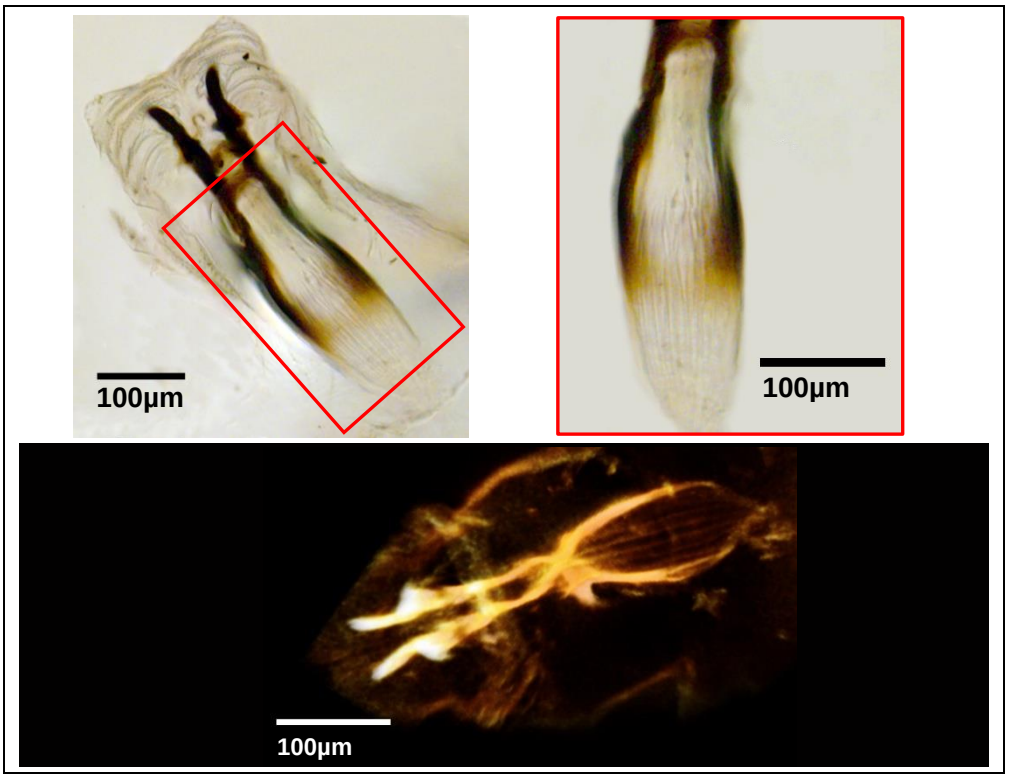

Fig. 19. *Drosophila melanogaster* sp. larva cephaloskeleton ventral view with light microscope (top) and confocal microscope (bottom)

The whiteflies that the *Acletoxenus* sp. larvae preyed on were identified as *Aleurotrachelus trachoides* (Back, 1912) based on the characters described in Fig. 20 (Martin, 1987; Walker, 2008). The COI sequences of the fourth instars, presented in Appendix, Table A2 (pg. A-4), were a 99% match to a sequence of *Aleurotrachelus trachoides* (Accession number KF059957) on GenBank. The samples sent overseas for verification were also identified as *Aleurotrachelus trachoides* (D. Barro, pers. comm.).

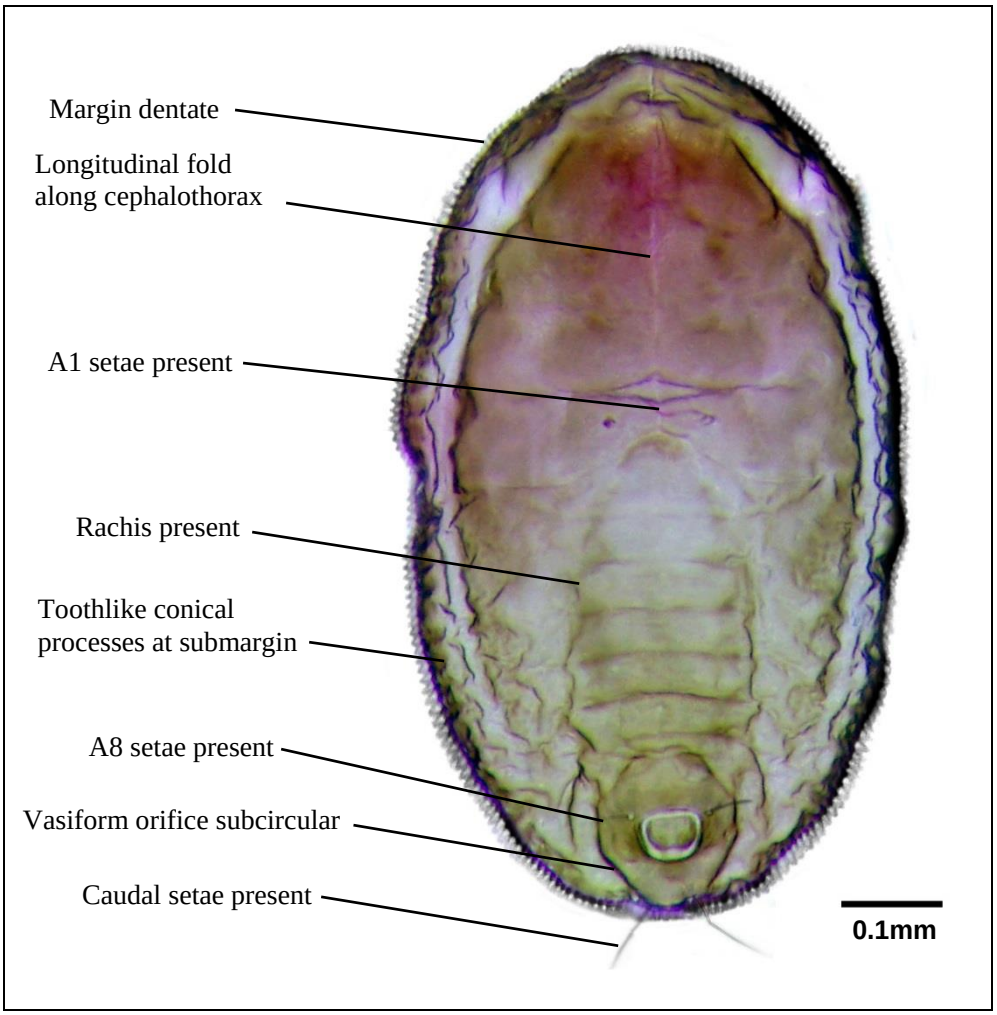

Fig. 20. Diagnostic features of the fourth instar of *Aleurotrachelus trachoides.*

The adult of *Acletoxenus* sp. has a proboscis which can be retracted in (Fig. 21, 22) or extended (Fig. 23 & 24). The proboscis consisted of the labrum which is

divided into two sponge-like labellar lobes (labella) at the tip (Fig. 24) (Colless & McAlpine, 1991). Each labellar lobe contained six grooves, known as the pseudotrachea that suggested capillary function (Fig. 24) (Elzinga & Broce, 1986). Two maxillary palps were attached to the labium and no mandibles were present (Fig. 23 & 24). Field and ex-situ observations also did not show any predation by the adults on the whitefly instars.

Fig. 21. Adult Acletoxenus sp. lateral view with proboscis folded in.

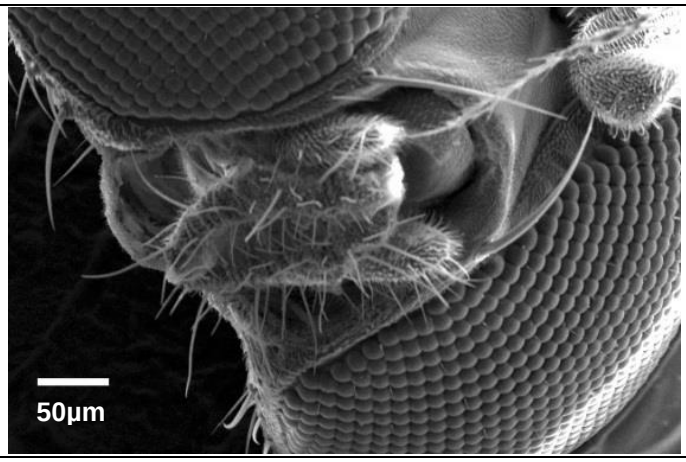

Fig. 22. SEM of *Acletoxenus* sp. adult head with proboscis folded in.

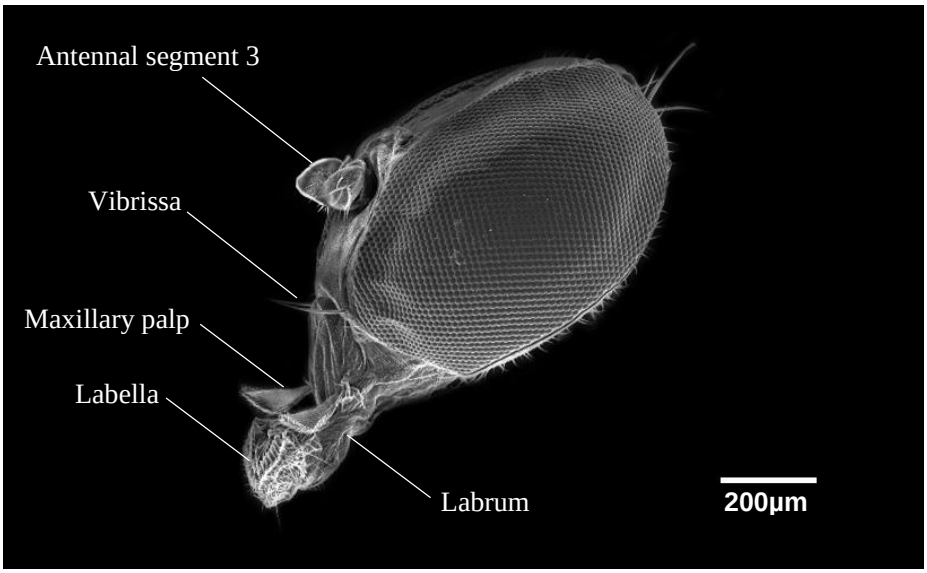

Fig. 23. SEM of *Acletoxenus* sp. adult head with proboscis extended out.

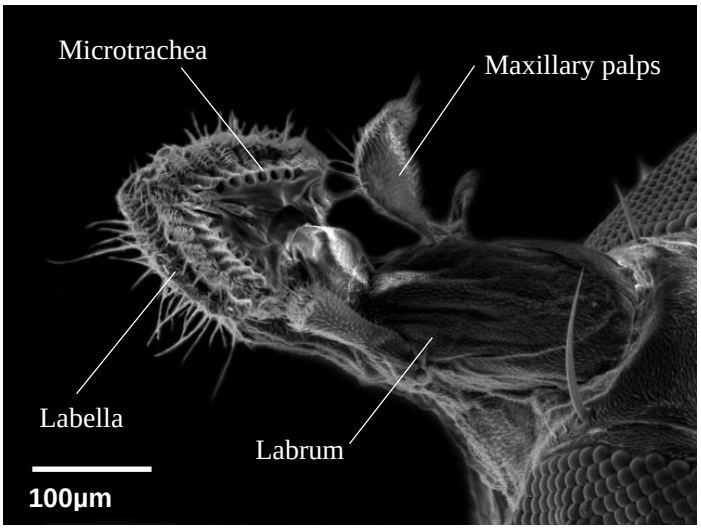

Fig. 24. SEM dorsal view of the *Acletoxenus* sp. adult mouthpart extended out

2.3.3 Natural History of *Acletoxenus* sp.

Acletoxenus sp. undergoes complete metamorphosis, having four life cycle stages: egg, larva, pupa and adult (Fig. 25). The mean time *Acletoxenus* sp. took to complete its life cycle was 24.1 days (Fig. 26). The 95% confidence interval of the time an egg (n = 86) took to hatch was (3.24, 3.73) days. The 95% confidence interval of the time spent as a larva (n = 88) was (11.8, 12.97) days.

The 95% confidence interval of the developmental time as a pupa (n = 128) was (8.14, 8.97) days. The 95% confidence interval of the lifespan of an adult (n = 31) was (10.3, 13.8) days.

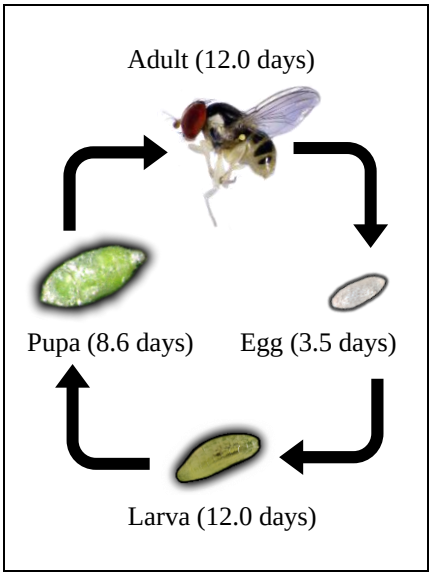

Fig. 25. Life cycle of *Acletoxenus* sp.

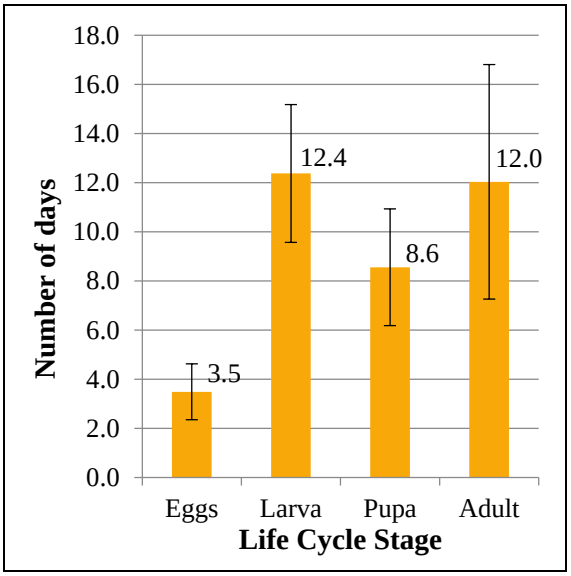

Fig. 26. Mean duration of each stage (error bars indicating standard deviation).

Females were observed to oviposit on leaves with early stages of whitefly instars in the morning and afternoon. The eggs were laid singly and the number of eggs oviposited on one leaf varied from one to four. They were approximately 0.45mm in length, 0.2mm in width and white in colour with somewhat indistinct reticulate markings. The eggs were attached strongly to the abaxial surface of the leaves in the horizontal position (Fig. 27).

Fig. 27. *Acletoxenus* sp. egg (left) found next to whitefly first instars (right)

The larvae were acephalic, cylindrical and tapered at the cranium. The larvae of the last instar were approximately 3-4 mm long in length and was widest along the middle of about 1mm (Fig. 28). The weakly sclerotized cuticle was thin and translucent, revealing cream coloured fat bodies in younger instars and green fat bodies in older instars (Fig. 28). The mouth hooks and the H-piece of the cephaloskeleton were the only visibly sclerotized organs (Fig.16).

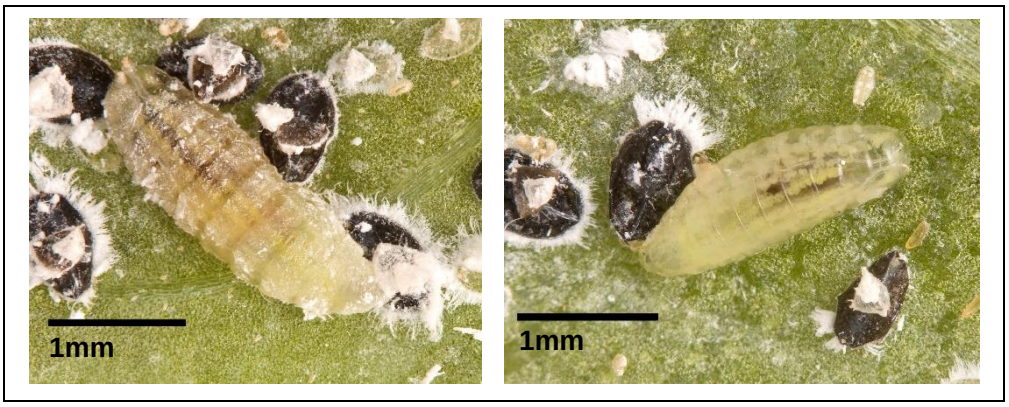

Fig. 28. Cream and green coloured larvae of *Acletoxenus* sp.

The body consisted of the pseudocephalon, three thoracic segments and eight abdominal segments (Fig. 29). The "head" or pseudocephalon where the sensory organs mouth and mouth hooks were found, was visibly retracted into the first thoracic segment (Fig. 15, 29 and 30). A pair of dorsolateral prothoracic spiracles were visible on the first thoracic segment, which were not seen in the other remaining segments (Fig. 29, 30 & 31). The second and third thoracic segments were similar with a general cylindrical shape (Fig. 29). Likewise, the abdominal segments were similar except that the ventral side seemed to be further sub-divided on the ventral surface (Fig. 29). The eighth abdominal segment (anal division) has a posterior pair of spiracles at its apex and a pair of pad-like anal organs on the ventral side (Fig. 32 & 33).

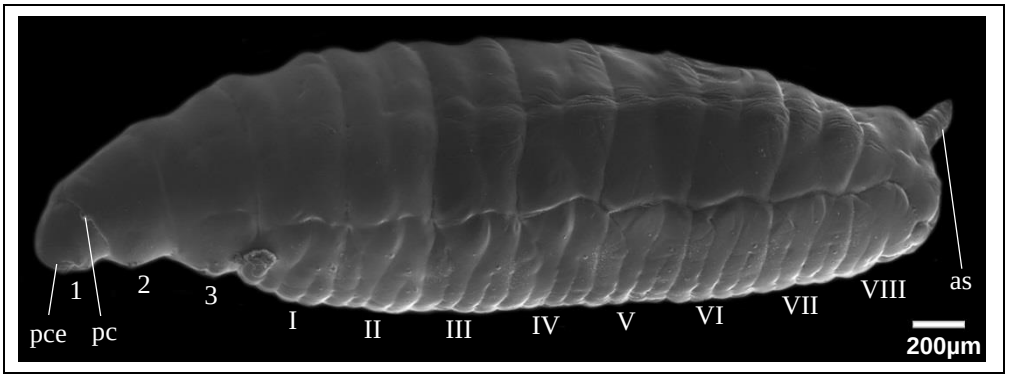

Fig. 29. Acletoxenus sp. larva lateral view. Abbreviations: 1-3, thoracic segments; I-VIII, abdominal segments (VIII = anal division); as, abdominal spiracles; pce, pseudocephalon; ps, prothoracic spiracles.

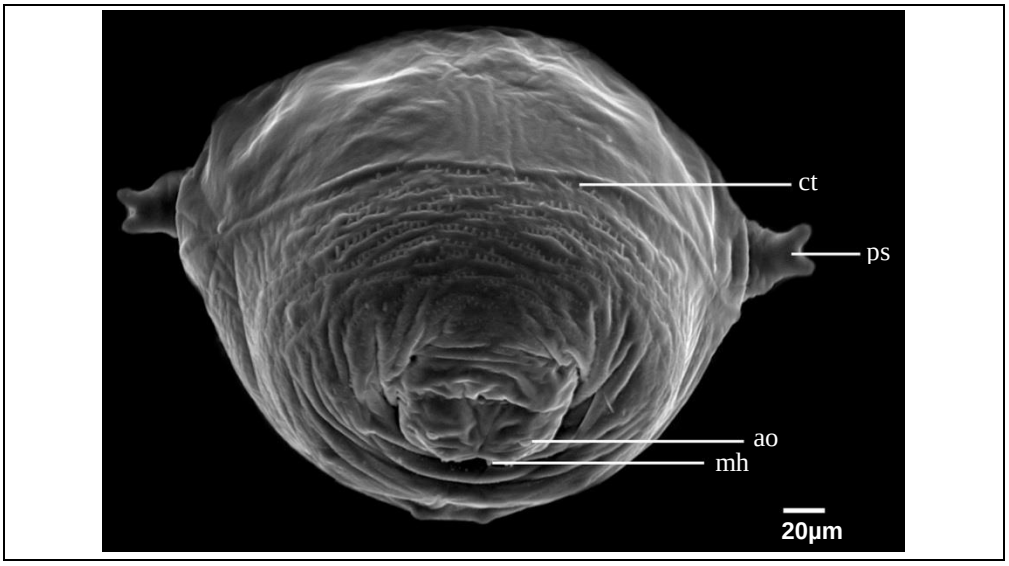

Fig. 30. *Acletoxenus* sp. larva front view of first thoracic segment and pseudocephalon. Abbreviations: ct, dendrite; ps, prothoracic spiracles; ao, antennal organ; mh, mouth hooks.

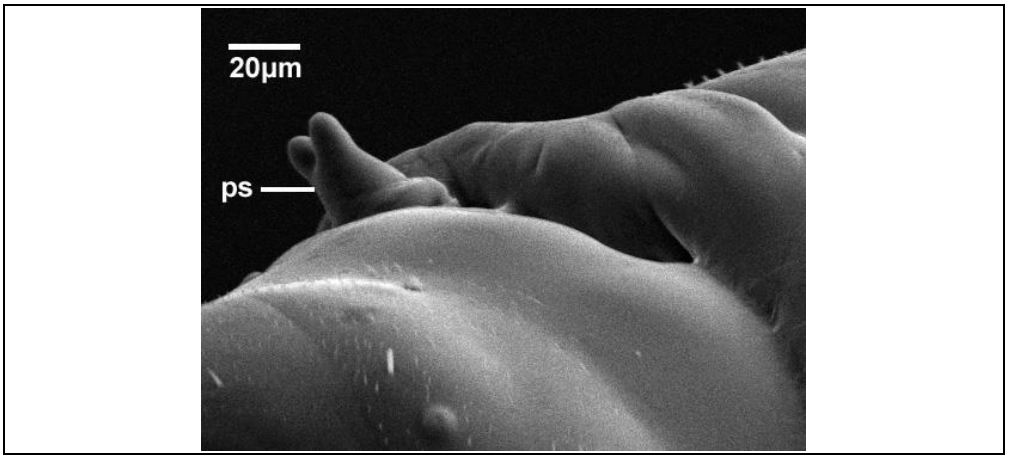

Fig. 31. *Acletoxenus* sp. larva prothoracic spiracles (ps).

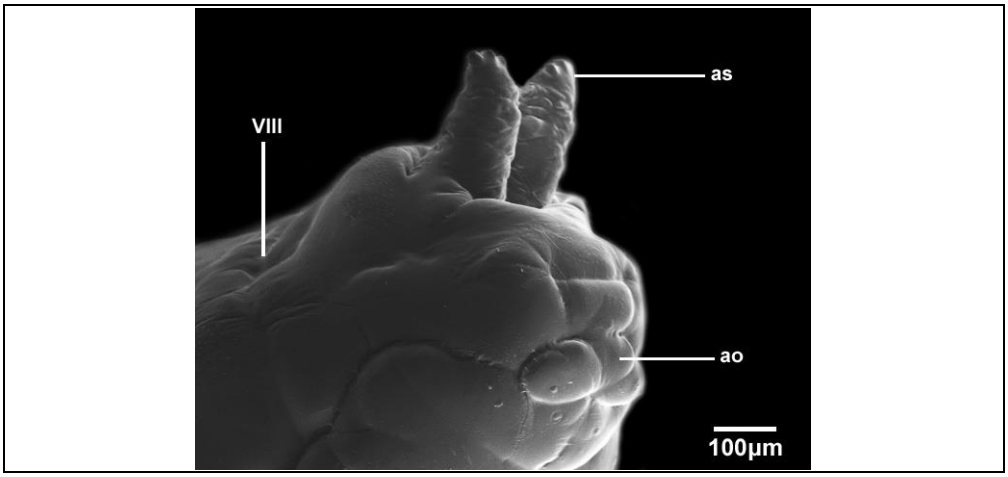

Fig. 32. *Acletoxenus* sp. larva anal division (abdominal segment VIII). Abbreviations: as, abdominal spiracles anus; ao, anal organ.

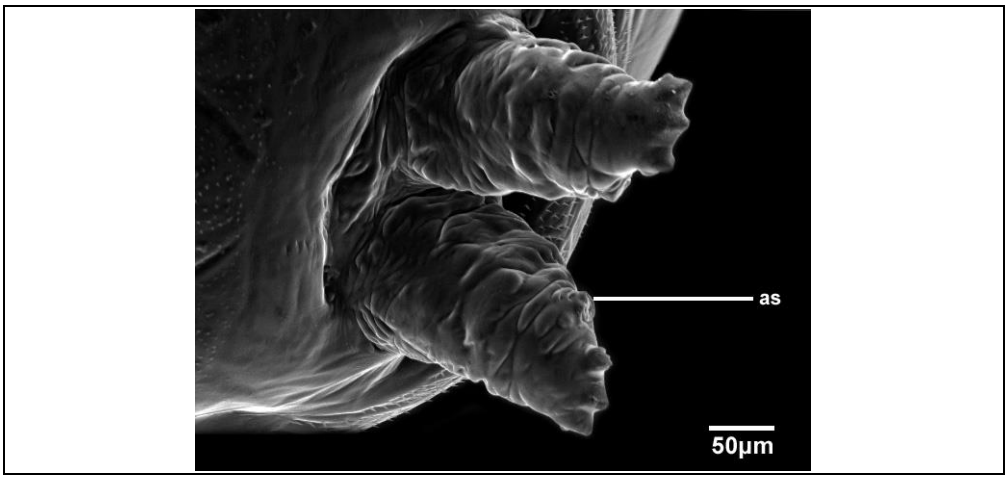

Fig. 33. *Acletoxenus* sp. larva abdominal spiracles (as).

The pupa of Acletoxenus sp. is cylindrical in shape tapering at the ends. It was about 3.3mm in length and 1.3mm at its widest width (Fig. 35 & 36). The ventral surface was flat and attached strongly to the leaf surface (Clausen & Berry, 1932). When the pupae were dislodged from the leaves and placed on moist tissue, glue was secreted by the pupae to re-attach themselves to the new surface. The dorsal surface of the pupa has two abdominal spiracles at the posterior end and a visible "lid" on the anterior end (Fig. 7 & 34). The integument was translucent and revealed the green colour of the body and red eyes that characterize the adults at later stages (Fig. 7 & 35). The larva pupated within

the last larval skin and was thus still covered by whitefly eggs, dead instars and wax (Fig. 36). The adults emerged by breaking open the distinct lid at the anterior end, leaving behind a translucent white puparium (Fig. 8 & 37) (Clausen & Berry, 1932).

Fig. 34. *Acletoxenus* sp. pupa with wax whitefly wax, eggs and dead instars removed.

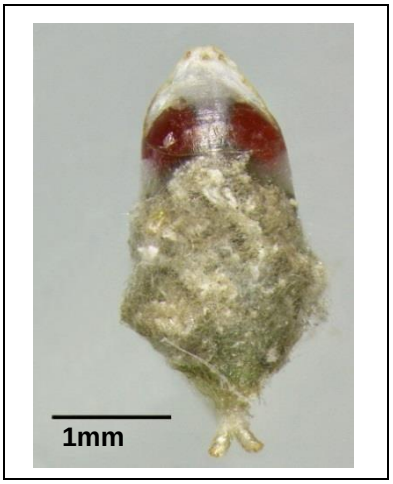

Fig. 35. Late stage *Acletoxenus* sp. pupa with red eyes.

Fig. 36. *Acletoxenus* sp. pupa covered in whitefly wax, eggs and dead instars (right) and with whitefly wax, eggs and instars removed (left).

Fig. 37. *Acletoxenus* sp. emergence from anterior end of pupa.

The mean monthly rate of parasitism was 43.3% with a standard error of 3.66% (Fig. 38). The highest rate of parasitism was in June while July saw a decrease in both the number of *Acletoxenus* sp. that successfully became adults and in the rate of parasitism. The parasitoid was identified as *Pachyneuron leucopiscida* Mani, 1939, based on the characteristics in Fig. 39 (Gupta *et al.*, 2013; Noyes, 2015; Sureshan & Narendran, 2000). The image sent for verification were also identified to the genera level of *Pachyneuron* (H. Bauron pers. comm.). The COI sequences of the parasitoid is provided in the Appendix, Table A3 (pg. A-4). Unfortunately, there was no sequence for *Pachyneuron leucopiscida* online for comparison. When the COI sequences were blasted to NCBI GenBank, the closest match was 89% to *Nasonia longicornis*. This species and *Pachyneuron leucopiscida* both belong to Pteromalinae.

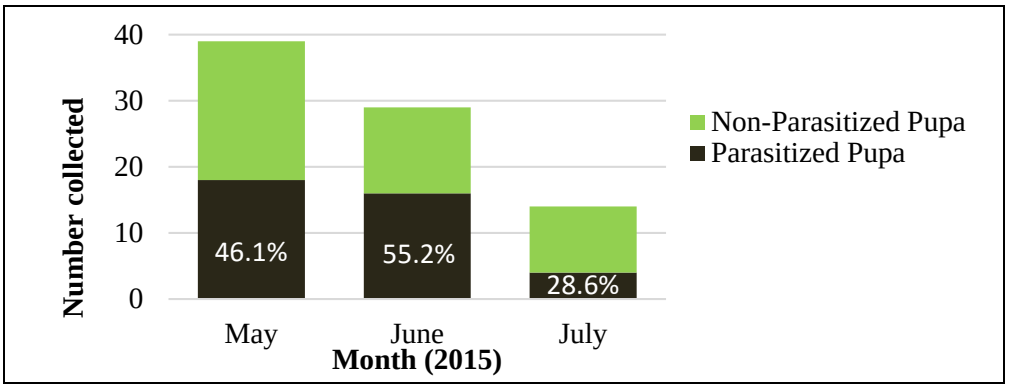

Fig. 38. Number of empty puparium collected.

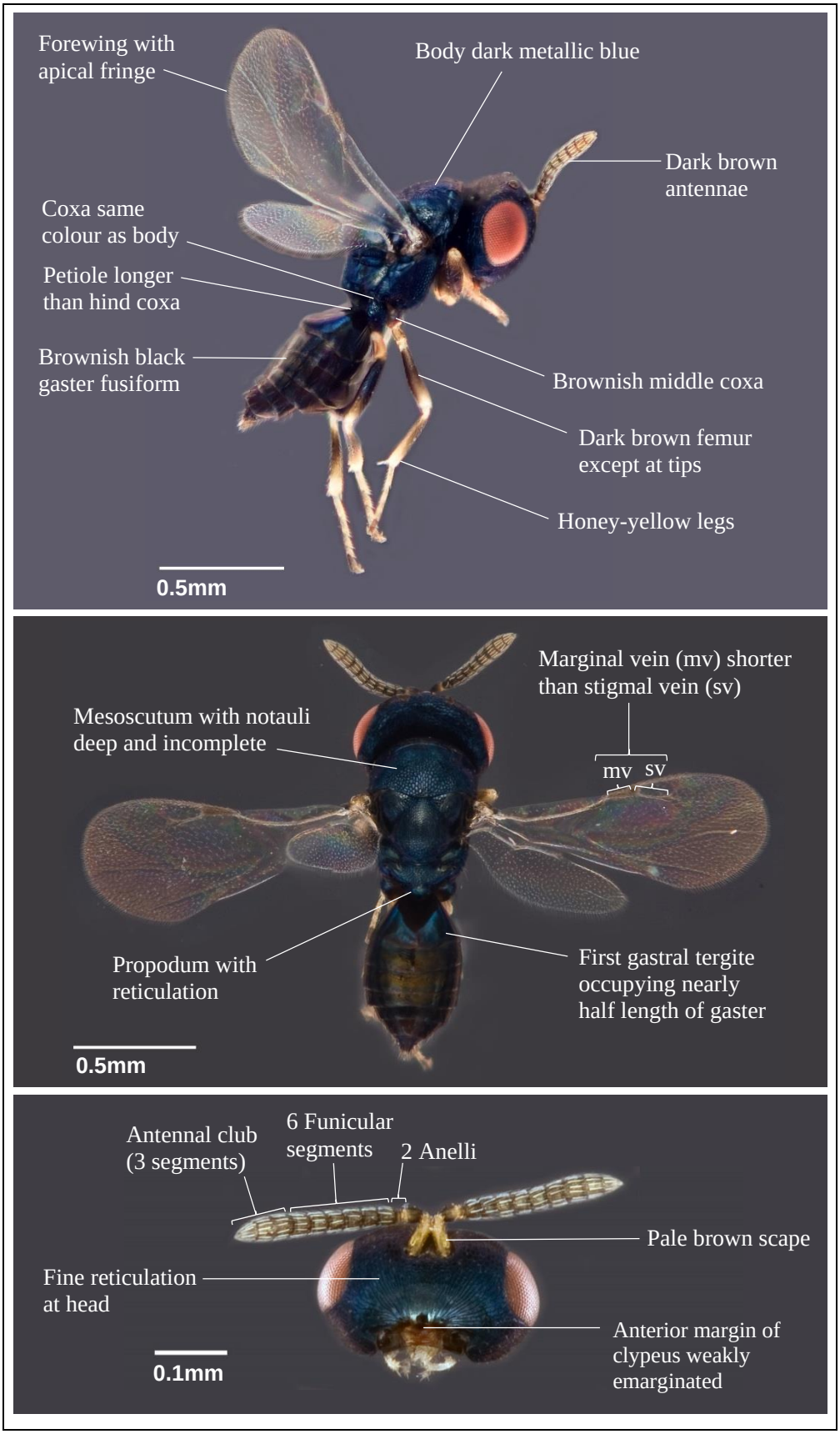

Fig. 39. Diagnostic features of *Pachyneuron leucopiscida.*

2.4 Discussion

2.4.1 The Identity of the Singapore *Acletoxenus* sp.

Most *Acletoxenus* species are identified based on the morphology of the mesonotum. However, this character system fails for the flies found in Singapore. This is because the species found in Singapore has morphological features known in three of the four known species: *Acletoxenus formosus*, *Acletoxenus indicus* and *Acletoxenus quadristriatus* (Fig. 10, 11 & 12). Since the DNA sequences of 12 individuals, representing all morphotypes and both genders, were almost identical with a low pairwise distance of 0.06%, it is very likely that these individuals belong to the same species. This implies that the morphology of the mesonotum is variable and should not be used for distinguishing species of *Acletoxenus.*

The original descriptions for *Acletoxenus indicus* were based on female specimens. This was a major obstacle in deciphering the species found in Singapore as gender and morphotypes were found to be significantly dependant by Fisher's exact probability test. As no females of *Acletoxenus quadristriatus* morphotype were found, there were non-significant differences between *Acletoxenus quadristriatus* and *Acletoxenus formosus* morphotype in addition to *Acletoxenus quadristriatus* and *Acletoxenus indicus* morphotype. As the differences between males of the three morphotypes were non-significant, it indicated that the proportion of males of each morphotype were similar. However, the females had a significant difference between *Acletoxenus formosus* morphotype and *Acletoxenus indicus* morphotype. This suggested that one would have a higher chance to find an *Acletoxenus formosus* morphotype

in females. Hence, if only a small number of females were caught and used for description, there would be a high chance for an incorrect species circumscription, i.e., if the description of the Singapore *Acletoxenus* sp. was based on the dominant morphology in females (entirely black mesonotum), the variability within the population would be overlooked.

The different morphotypes of the Singapore *Acletoxenus* sp. all had a dorsocentral black mark on the fourth tergite and a much smaller mark of similar form on the fifth tergite, which was similar to what was described for *Acletoxenus indicus*. However, the morphology of the abdomen has been variable in *Acletoxenus formosus* (Bock, 1982; Collin, 1902; Malloch, 1929) and since the descriptions of the abdomen were from one source only in *Acletoxenus indicus* (Malloch, 1929) and *Acletoxenus quadristriatus* (S. McEvey, pers. comm.), one would be cautioned against using the abdomen morphology to differentiate species.

One remaining morphological feature that had been used in identification keys was the length of the orbital setae. Malloch (1929) distinguished *Acletoxenus indicus* from *Acletoxenus formosus* and *Acletoxenus meijerei* by proclinate orbital setae that were noticeably shorter than the anterior reclinate setae. Bock (1982) translated Duda's (1936) description of *Acletoxenus quadristriatus* with anterior reclinate setae that were stronger than that of the proclinate orbitals. Since the Singapore species had proclinate orbital setae that were noticeably shorter than the anterior reclinate ones, this ruled out *Acletoxenus formosus* and *Acletoxenus meijerei*, narrowing down the species choice to *Acletoxenus indicus*, *Acletoxenus quadristriatus*, or a new species (Fig. 7). The pairwise distance of

11.14% when compared to a voucher specimen of *Acletoxenus formosus* supported that Singapore species was not it, although its defining morphology of an entirely black mesonotum was present in the Singapore species.

Finally, the pairwise distance of 1.69% when compared to the voucher specimen of *Acletoxenus indicus* indicates that flies from Singapore are likely to be *Acletoxenus indicus* given that most species differ by COI distances >2%. However, if a lower threshold of 1% was used for species limitation, then the species in Singapore could be considered a separate species. Thus, based on the currently available evidence, one should treat the species as *Acletoxenus* cf. *indicus*.

2.4.2 Are *Acletoxenus* sp. Predators?

The external morphology of *Acletoxenus* cf. *indicus* larvae was largely similar to that of other Cyclorrhapha larvae (Fig. 29). However, as a predator, *Acletoxenus* cf. *indicus* lacked a well-developed facial mask, which is present in saprophagous Cyclorrhapha larvae for rasping and directing bacteria into their mouths (Fig. 15) (Courtney *et al.*, 2000; Dowding, 1967; Roberts, 1971). *Acletoxenus* cf. *indicus* also lacks a filter apparatus that is used by saprophagous larvae to filter bacteria for ingestion (Fig. 18). Because the filter is missing, the cephaloskeleton of *Acletoxenus* cf. *indicus* is less sclerotized compared to that of *Drosophila melanogaster* which has musculature that attaches to the cephaloskeleton for this purpose (Fig 16).

Although Clausen & Berry (1932) recorded *Acletoxenus indicus* to be inactive and never leaving from the leaf upon which the egg was laid, the species found in Singapore did move to other leaves in order to locate prey, albeit at very slow

speeds. *Acletoxenus formosus* larvae have been estimated to feed on 30 to 40 whitefly puparia during their development, which might be similar to that for the Singapore species. (Pelov & Trenchev, 1973). The adults however were probably not predators as they lacked mandibles (Fig. 23 & 24). The sponging mouthparts also suggested that the adults rely on sucking up fluids for nutrition (Fig. 24).

The prey of the *Acletoxenus* cf. *indicus*, *Aleurotrachelus trachoides,* is a major pest on kava and capsicum in the federal states of Micronesia (PestNet, 2011). Thus, one could propose *Acletoxenus* cf. *indicus* as a potential biological control agent. However, attempts in the past to use *Acletoxenus indicus* (Clausen & Berry, 1932) and *Acletoxenus formosus* (Vayssière, 1953) in biological control have failed. Although the reasons for the failure were never fully investigated, extensive parasitism by Hymenoptera was suggested as a contributing factor (Clausen & Berry, 1932; Mentzelos, 1967; Pelov & Trenchev, 1973). This study also found *Acletoxenus* cf. *indicus* to be parasitized by *Pachyneuron leucopiscida* at a high rate (mean = 43.3%) which may have contributed to the large drop in population size of *Acletoxenus* cf. *indicus* which was observed in the last few months of my study (Fig. 38). Thus, I conclude that *Acletoxenus* cf. *indicus* shows low potential as a biological control agent and that high parasitism rates may be one of the main reasons.

2.4.3 Natural History of *Acletoxenus* cf. *indicus*

The mean time for *Acletoxenus* cf. *indicus* to complete its life cycle is 24.1 days which is similar to the life cycle duration of *Acletoxenus formosus* that is found in Europe of around 12 (Frauenfeld, 1868) to 27 days (Pelov & Trenchev, 1973).

The approximately 0.45mm long eggs were slightly larger than the 0.4mm length recorded for *Acletoxenus indicus* by Clausen and Berry (1932). Although Clausen and Berry (1932) mentioned that oviposition occurs during midday, *Acletoxenus* cf. *indicus* was observed to oviposit in the morning as well. Eggs are oviposited on leaves with earlier stages of whitefly instars which provides the newly hatched *Acletoxenus* cf. *indicus* larvae more manageable prey of smaller size. In addition, eggs were laid singly, similar to the observations of Clausen and Berry (1932).

The larvae secreted a mucus which caused whitefly wax, egg and puparium to become stuck on its body (Fig. 14) (Ashburner, 1981; Clausen & Berry, 1932). Possible functions for this could be to help camouflage the larvae against predators, to provide an addition layer of protection over its thin sclerotized cuticle and to prevent desiccation/parasitism. *Acletoxenus* cf. *indicus* was more mobile in moving to other leaves, in comparison to *Acletoxenus indicus* larvae observed to never leave to another leaf (Clausen and Berry, 1932). This behaviour would allow *Acletoxenus* cf. *indicus* larvae to reduce intra-specific competition. During metamorphosis into pupae, the larvae would retain the same cuticle layer that has picked up the whitefly wax, egg and puparium (Fig. 36) and continue to serve the proposed functions.

The pupae constantly secreted a strong glue to keep themselves adhered to the ventral side of leaves. This allowed them to be able to adhere to a new surface even when they get dislodged during a big gust or thunderstorm which are common in tropical Singapore. Unfortunately, the pupae were highly susceptible to parasitoid attacks by *Pachyneuron leucopiscida*, which has been

recorded as a parasitoid on *Acletoxenus indicus* (Noyes, 2015). When the parasitism rates increased, the population of *Acletoxenus* cf. *indicus* decreased over time (Fig, 38). By August, there was a big decrease in *Acletoxenus* cf. *indicus* such that the population could not recover and no new individuals were found. This terminated the experiment before other tests on feeding rates could be carried out.

Chapter 3. *Acletoxenus* cf. *indicus* Transcriptomics

3.1 Introduction

Although both species are classified as Drosophilidae, the proteins expressed between the entomophagous *Acletoxenus* cf. *indicus* and the saprophagous *Drosophila melanogaster* can be expected to be vastly different due to the differences in the occupied niche. For example, an entomophagous organism would have to express enzymes that digest chitin while a saprophagous organism would have to express proteins to help deal with extreme microbial stress (Altincicek & Vilcinskas, 2007). Thus, a comparative transcriptomic study was conducted on *Acletoxenus* cf. *indicus* larvae with comparisons against *Drosophila melanogaster* and Drosophilidae genes to discover what these differences in gene expression were that reflect the difference in their life history. The aims were to:

1. Discover *Acletoxenus* cf. *indicus* genes that were evolutionarily distinct from other drosophilids (divergent genes)
2. Discover the potential functions of those divergent genes that were highly expressed
3. Compare difference in expression levels of orthologous genes between *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* in order to find potential adaptations in expression levels

3.2 Material and Methods

3.2.1 RNA Extraction and Transcriptome Assembly

Total RNA was extracted from whole body tissue of *Acletoxenus* cf. *indicus* larvae using phenolic trizol reagent (Bogart & Andrews, 2006). Illumina paired-end sequencing libraries were then prepared with TruSeq Stranded mRNA Library Prep Kit before being sequenced on Illumina Genome Analyzer HiSeq 2500 platforms (Illumina, San Diego, CA, USA). CLC Genomics Workbench 6.5.1 (CLC bio, Aarhus, Denmark) was then used to assemble the transcriptome data. First, the transcriptomic reads were filtered by quality trimming using the quality score limit of 0.01 and ambiguities score of 2, removing any reads that were less than 20 base pairs long. *De Novo* assembly was then conducted using an optimized k-mer length and bubble size of 50, retrieving only contigs with a minimum length of 200 base pairs. The reads were then mapped back to the contigs in slow-contig mapping mode (mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 1.0). Thereafter, the assembled contigs were filtered for contigs that had an average coverage of more than 10, to remove the low quality contigs. The relative expression levels of each contig were then categorized by using the outlier formula on the read counts as followed (Agresti & Franklin, 2013):

(1) *Very low expression* $\leq Q_1 - 3.0IQR$

(2) $Q_1 - 3.0IQR \leq$ *low expression* $\leq Q_1 - 1.5IQR$

(3) $Q_1 - 1.5IQR \leq$ *moderate expression* $\leq Q_3 + 1.5IQR$

(4) $Q_3 + 1.5IQR \leq$ *high expression* $\leq Q_3 + 3.0IQR$

(5) *Very high expression* $\geq Q_3 + 3.0IQR$

3.2.2 Discovering Divergent Genes

Highly divergent genes in *Acletoxenus* cf. *indicus*, would consist of contigs that could not be easily matched to the known genes in a saprophagous drosophilid. Thus, one can first identify and eliminate easily matched contigs which are orthologous to genes in the saprophagous drosophilid. This was done by first creating a local database of nucleotide and amino acid (AA) annotations of *Drosophila melanogaster*'s genomes through downloading the data from NCBI GenBank (accession number NC_004354.4, NT_033779.5, NT_033778.4, NT_037436.4, NT_033777, NC_004353.4, NC_024512.1, NC_024511.2) (Benson *et al.*, 2013). A reciprocal blast was then performed between the *Drosophila melanogaster* annotated nucleotide dataset and the *Acletoxenus* cf. *indicus* contigs using the blastn algorithm with minimum similarity of $1e^{-5}$ (Altschul *et al.*, 1990). Pairs that matched in the reciprocal blast were considered orthologues. These orthologues were considered to have little evolutionary differences because they could be homologized easily based on fairly strict identity criteria. Thus, these orthologous contigs in the *Acletoxenus* cf. *indicus* contig dataset were removed with a custom python script.

As proteins are more conserved then nucleotide sequences, the same step can be repeated with AA translations in order to remove all *Acletoxenus* genes with low AA divergence. The contigs left over from the previous step were translated to amino acids in normal and reverse complement order with Virtual Ribosome dna2pep ver. 1.1, trying all reading frames and using ORF mode none with invertebrate mitochondrial and standard genetic codes (Wernersson, 2006). The translated contigs were then blasted reciprocally to the *Drosophila*

melanogaster annotated amino acid dataset using the blastp algorithm with minimum similarity of $1e^{-5}$ to locate for orthologues that were missed out during blastn (Altschul *et al.*, 1990). The contigs that had a match were once again removed with a custom python script because they would have few evolutionary relevant differences.

The remaining contigs that did not get matched in the blastn and blastp step could either be very divergent genes or genes that were lab sequencing contaminations. Thus, a quality control step was used to remove contigs with non-drosophilid matches. The remaining contigs that did not match in the blastn and blastp steps were blasted with MEGAblast algorithm with minimum similarity of $1e^{-5}$ to the online NCBI GenBank database (Altschul *et al.*, 1990). Contigs that matched to a blast hit that was not Drosophilidae were removed with a custom python script. After removing these contigs, the resulting contigs represented divergent genes in *Acletoxenus* cf. *indicus* that could not be easily matched to those of *Drosophila melanogaster*, a saprophagous drosophilid.

To get an idea of the function of the list of contigs that represented divergent *Acletoxenus* cf. *indicus* genes, a more relaxed blast setting was used in another round of blast. This included blasting against all Drosophilidae genes in the NCBI database instead of just the genome of *Drosophila melanogaster* as well as using the blastn algorithm with minimum similarity of $1e^{-5}$ with a smaller word size of 7 (Altschul *et al.*, 1990).

3.2.3. Comparing to the Drosophilidae Genome

A subset of the final result from Section 3.2.2 was created by picking contigs that had relatively high expressions (read counts above 2879) for further

analyses. This subset of 26 contigs was blasted to the NCBI Drosophilidae database with blastn algorithm with a more relaxed criteria of minimum similarity of $1e^{-2}$ and word size 7 (Altschul *et al.*, 1990). Next, the aligned region of a blast hit match which had the highest identification percentage was extracted. This region was then checked that it was part of the coding strand by aligning to the coding sequence entry of the blast hit with MAFFT ver. 7 on automatic settings (Katoh & Standley, 2013). If the sequence was not a 100% match to a part of the coding sequence, it indicated that the current blast hit was a non-coding sequence. Thus, the next blast hit match with the highest identification percentage was chosen until the criteria was met. MEGA6 was then used to translate the aligned sequences (Tamura *et al.*, 2013). The correct reading frame was determined by checking if the translated coding strand was 100% identical to the translation from the NCBI database. Thereafter, the translated coding strand sequence was removed and the pairwise distances between the translated query and match hit was calculated with MEGA6 (Tamura *et al.*, 2013). This pairwise distance would indicate the difference between the proteins produced by the *Acletoxenus* cf. *indicus* and the species from the blast hit. The functions of each contigs were then determined by looking at the ontology of the blast hit through referencing the database of FlyBase (dos Santos *et al.*, 2015), UniProt (TheUniProtConsortium, 2015) and InterPro (Mitchell *et al.*, 2015).

3.2.4. Comparing to *Drosophila melanogaster* Larval Transcriptomes

To compare the gene expression levels between the larval stage of *Acletoxenus* cf. *indicus* and *Drosophila melanogaster*, transcriptomic data of *Drosophila melanogaster* larvae were downloaded from the NCBI Sequence Read Archive

(id: SRR1197324, SRR1197308, SRR1197312 and SRR1197307) to create a dataset for comparison. After performing the same quality trimming steps, a subset of the downloaded *Drosophila melanogaster* dataset was created to correspond to the same number of reads as the *Acletoxenus* cf. *indicus* dataset, using the extract subset tool on CLC Genomics Workbench 6.5.1. The subset was then assembled with the same settings as the *Acletoxenus* cf. *indicus* dataset.

A reciprocal blast was then performed between the *Acletoxenus* cf. *indicus* contigs and *Drosophila melanogaster* contigs using the blastn algorithm with minimum similarity of $1e^{-5}$ with word size 7 (Altschul *et al.*, 1990). The resulting pairs of contigs were then blasted to the NCBI Drosophilidae database using blastn algorithm with minimum similarity of $1e^{-5}$ with word size 7 to get an annotation. Only the pairs of contigs that had the same results in the blast to the NCBI Drosophilidae database were considered to be orthologues. To determine if the orthologues were coding regions, the aligned regions between the *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* orthologues were extracted. These regions were then checked that if it were part of the coding strand by aligning to the coding sequence entry of the blast hit result of the NCBI Drosophilidae database with MAFFT ver. 7 on automatic settings (Katoh & Standley, 2013). If the *Drosophila melanogaster* contig sequence was not a 100% match to a part of the coding sequence, the next common blast hit match with the highest identification percentage was chosen until the criteria was met. If none of the common blast hits were a match, the orthologues were deemed to be non-coding regions. MEGA6 was then used to translate the aligned sequences and the translated coding strand sequence (Tamura *et al.*, 2013). The correct reading frame was determined by checking if the translated *Drosophila*

melanogaster contig was 100% identical to the translation from the NCBI database. Thereafter, the pairwise distances between the translated query and match hit was calculated with MEGA6 (Tamura *et al.*, 2013).

3.3 Results

3.3.1 Transcriptome Assembly and Divergent Gene Discovery

Out of the 48,474,776 reads (250 bp PE), 35807 and 48122 contigs were assembled from the *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* transcriptomic data respectively. After removing low quality contigs that had average coverage of less than ten, 21488 and 19620 contigs remained for *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* respectively. The relative expression levels of each contigs using the outlier formulas were determined by the read count values listed in Table 1. A large majority (87.6%) of the *Acletoxenus* cf. *indicus* contigs were expressed at relatively moderate levels with the remaining 4.5% and 7.9% being expressed relatively at high and very high levels (Fig. 40). In contrast, 24.9% of *Drosophila melanogaster* contigs were relatively lowly expressed, 62.1% was relatively moderately expressed, 4.3% was relatively highly expressed and 8.7% was relatively very highly expressed (Fig. 41).

Table 1. Relative gene expression levels by read counts.

Relative Gene Expression Level	*Acletoxenus* cf. *indicus* Read Counts	*Drosophila melanogaster* Read Counts
Very Low	NA	Less than 0
Low	Less than 0	Between 0 to 78
Moderate	Less than 1829	Between 78 to 985
High	Between 1829 to 2879	Between 985 to 1530
Very High	More than 2879	More than 1530

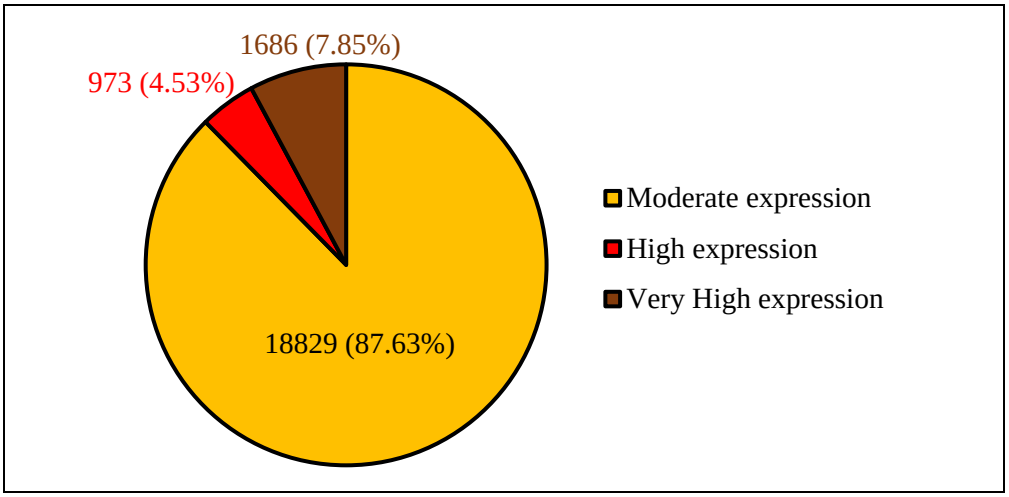

Fig. 40. Relative expression levels of *Acletoxenus* cf. *indicus* contigs.

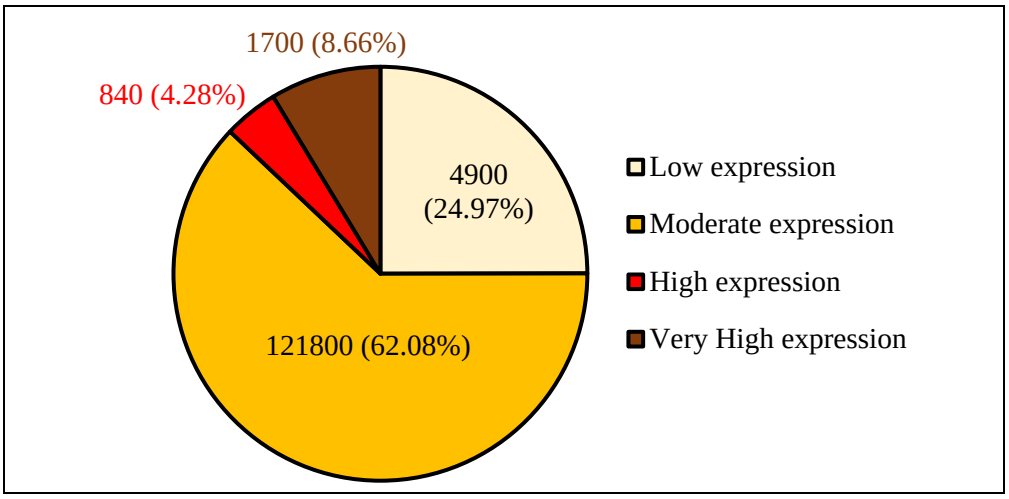

Fig. 41. Relative expression levels of *Drosophila melanogaster* contigs.

During the reciprocal nucleotide blast against the database of 17814 gene annotations to locate for orthologues, 10922 contigs of *Acletoxenus* cf. *indicus* returned with a hit. After removing those contigs, the remaining 10566 translated contigs resulted in 2802 orthologous matches with the reciprocal protein blast of 30324 protein annotations. The remaining 7764 contigs had 185 contigs that had a blast hit to a species that was not Drosophilidae with MEGAblast. These 185 contigs included bacteria, ants, moths, tomatoes and whiteflies, which were species that were found in the same locality where *Acletoxenus* cf. *indicus* was collected. Additional contigs were from model

organisms which might have been introduced from the Illumina machines (Appendix Table A4, pg. A-5). After removing these contigs, there were 7579 contigs that represented genes that were divergent from other drosophilids. These 7579 contigs had 104 contigs (1.37%) that had relatively very high expression levels, 44 contigs (0.58%) that had relatively high expression levels and the remainder 7431 (98.5%) that had relatively moderate expression levels (Fig. 42). Out of these 7579 contigs, 1551 contigs had matches during the relaxed blast to the NCBI Drosophilidae database and 6028 contigs that had no matches.

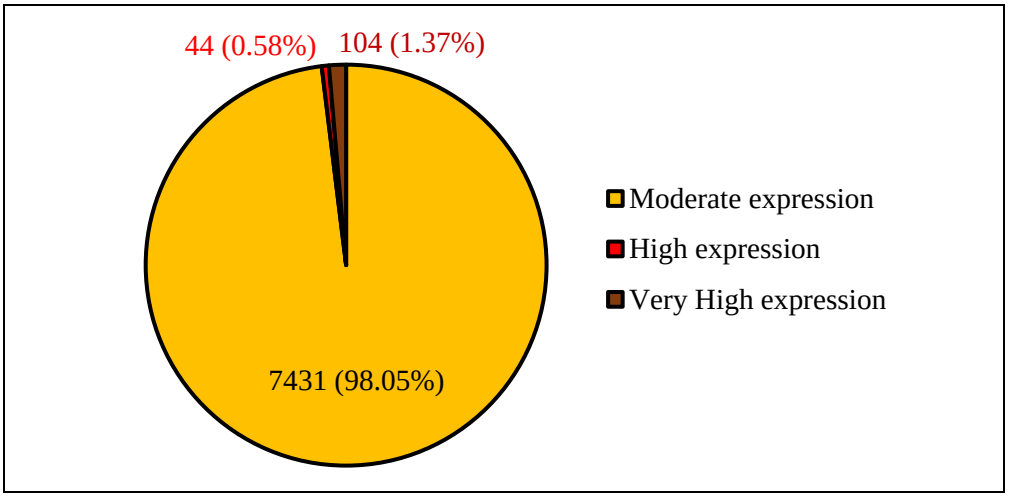

Fig. 42. Relative expression levels of the divergent contigs of *Acletoxenus* cf. *indicus*

3.3.2 Comparison of Relatively Highly Expressed Contigs in *Acletoxenus* cf. *indicus* to The Drosophilidae Genome

Out of the 1551 contigs that had a match during the relaxed blast to the NCBI Drosophilidae database, there were 26 contigs that were very highly expressed (with read counts more than 2879). The summary of the possible role of each gene/contig was listed in Table 2. The alignments between the contigs and their blast hits are presented in Appendix, Fig. A1 (pg. A-6). The highest in read

counts was Contig 122 that matched to Dmel\CG8260, which unfortunately has molecular and biological processes that have not been been discovered. However, the protein translated from this gene contained the BTB/POZ domain which is found in proteins that had many cellular functions, including transcription regulation, cytoskeleton dynamics, ion channel assembly and gating, and ubiquitination of target proteins (Stogios *et al.*, 2005). The second highest in average coverage was contig 62, that matched to Dmoj\GI12346. This gene is involved in the structural constituent process of binding chitin in insect cuticles (Rebers & Willis, 2001). The third highest in average coverage was contig 18 that matched to Dwil\GK18671. Although the molecular and biological processes that it are involved are unknown, the protein produced by Dwil\GK18671 contains the haemocyanin/hexamerin domain. Haemocyanin and hexamerin are proteins found in the haemolymph of invertebrates. Haemocyanin transports oxygen while hexamerin serves as a store of amino acids for synthesis of adult proteins (Beintema *et al.*, 1994).

Contig 580 was matched to Dwil\GK10515. The protein it produces contained the chitin binding peritrophin-A domain and is involved in chitin binding in chitin metabolic processes (Elvin *et al.*, 1996). Contig 285, which matched to Dgri\GH16331, is involved in the negative regulation of target of rapamycin (TOR) signalling and positive regulation of guanosine triphosphate hydrolase (GTPase) activity (FlyBase Curators *et al.*, 2004). Contig 228 was matched to Dper\GL14534. Although the molecular and biological processes of the gene are unknown, the protein it produces contained the PWWP domain. This domain is present in nuclear proteins and played a role in cell growth and differentiation involving protein-protein interactions (Stec *et al.*, 2000).

Contig 506 was matched to Dvir\Obp56d. This gene produces an odorant-binding protein with the PBP/GOBP domain that is associated with pheromone-sensitive neurons and general-odorant binding proteins (Vogt *et al.,* 1991). Contig 851 was matched to Dpse\GA28531 and Dper\GL17904 with the same identity score. Both genes are involved in unknown molecular and biological processes but produces proteins with the GYR and YLP motif. These motifs are substrates for tyrosine kinases and have a role in cuticle assembly (Cornman, 2010). Contig 108 was matched to Dmel\Muc26B. The protein produced by this gene functions as an extracellular matrix structural constituent and is involved in chitin metabolic processes and neuron projection morphogenesis (FlyBase Curators *et al.*, 2004; Syed *et al.*, 2008).

Contig 3038 was matched to the mitochondrial NADH-ubiquinone oxidoreductase chain 2 (ND2) gene of *Phortica* sp. ND2 is involved in NADH dehydrogenase (ubiquinone) activity and ATP synthesis. Contig 16 was matched to the larva glue protein (lgp1) gene in *Drosophila virilis*. This gene is responsible for the glue protein that adheres pre-pupa and subsequently pupa to a substrate for the duration of pupation (Lanio *et al.,* 1994). Contig 1468 was matched to Dmoj\GI18458 that produces proteins with the peptidase C1A, proteinase inhibitor I29 and cathepsin propeptide domain. It is involved in cysteine-type peptidase activity (FlyBase Curators *et al.*, 2004).

Contig 2887 was matched to Dvir\GJ24617. It translated into a key enzyme for alkaloid biosynthesis and is involved in strictosidine synthase activity (Maresh *et al.*, 2008). Contig 2788 (Dgri\GH12332) and 4057 (Dgri\GH19937) had gene identifications whose ontology has not been discovered (Burman *et al.*, 2014;

FlyBase *et al.*, 2004). Contig 1463 was matched to Dwil\GK20435 The protein produced by this gene contains the ubiquitin-conjugating enzyme active site and thus may be involved in selective degradation of cellular proteins (Hershko, 1991).

Contig 4330 was matched to Dvir\GJ21176 which produces proteins of the mitochondrial import receptor subunit TOM7 family. It is involved in protein import into the mitochondrial matrix (Jänsch *et al.*, 1998). Contigs 6164 and contig 6576 was matched to Dwil\GK13511 and Dvir\GJ24113 respectively. Both matches has protein domains that are involved in nucleic acid and nucleotide binding (FlyBase Curators *et al.*, 2004). Contigs 4109 and 2852 were matched to Dvir\GJ15528 and Dwil\Gk11746 correspondingly, whose ontologies have not been determined. Contigs 4428, 1890. 4117 (Dmel\CR44643), 766 and 3161 had only noncoding regions that matched in the blast hits, so their possible role could not be determined. The alignment length of most of the aligned regions were less than 50 amino acids long except in contig 18, 506 and 4330 (Table 2). The pairwise distance between the contigs and their blast hits of more than 11% also indicated that the proteins produced were different (Table 2).

Table 2. Summary of relatively very highly expressed divergent contig blast hit matches.

Contig No.	Read Count	Gene Match	AA alignment length	AA p-distance	Possible role
Contig 122	471473	Dmel\CG8260	17	33.60%	Housekeeping gene
Contig 62	317376	Dmoj\GI12346	16	20.80%	Binding chitin in cuticle
Contig 18	217804	Dwil\GK18671	66	34.00%	Oxygen transport, protein storage
Contig 580	55611	Dwil\GK10515	47	71.50%	Binding chitin in cuticle
Contig 285	32627	Dgri\GH16331	11	20.10%	Positive regulation of GTPase & negative regulation of TOR signalling
Contig 228	31566	Dper\GL14534	23	57.10%	Cell growth and differentiation
Contig 506	27503	Dvir\Obp56d	66	45.20%	Sense of smell
Contig 851	13306	Dpse\GA28531 & per\GL17904	25	22.30%	Target for tyrosine kinases; cuticle assembly
Contig 108	11713	Dmel\Muc26B	89	57.50%	Chitin metabolic process & neuron projection morphogenesis
Contig 3038	8276	Phortica\ND2	10	10.50%	ATP synthesis
Contig 16	8087	Dvir\lgp1	37	87.50%	Codes for glue so pupa can adhere to substrate
Contig 1468	7400	Dmoj\GI18458	44	58.30%	Cysteine-type peptidase activity
Contig 2887	5864	Dvir\GJ24617	26	36.80%	Alkaloid biosynthesis
Contig 2788	5413	Dgri\GH12332	19	74.70%	Unknown
Contig 4057	5038	Dgri\GH19937	20	11.10%	Unknown
Contig 1463	5025	Dwil\GK20435	22	52.60%	Cellular degradation of proteins
Contig 4330	4469	Dvir\GJ21176	54	22.80%	Protein import into mitochondria
Contig 4428	4417	Non-coding			
Contig 1890	4255	Non-coding			
Contig 4117	3906	Dmel\CR44643 (Non-coding)			
Contig 6164	3536	Dwil\GK13511	41	13.00%	Nucleic acid and nucleotide binding
Contig 6576	3159	Dvir\GJ24113	17	12.50%	Nucleic acid and nucleotide binding
Contig 766	2987	Non-coding			
Contig 4109	2981	Dvir\GJ15528	35	26.00%	unknown
Contig 2852	2963	Dwil\GK11746	49	75.60%	unknown
Contig 3161	2962	Non-coding			

3.3.3 Comparison of *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* Larval Transcriptomes

The 1551 *Acletoxenus* cf. *indicus* contigs had 117 orthologous matches to the 19620 *Drosophila melanogaster* contigs during the reciprocal blast. The final coding region and annotation check resulted in 12 annotated orthologous contig pairs (Table 3). The alignments of the orthologous contig pairs are presented in Appendix Fig. A2 (pg. A-7).

Acletoxenus cf. *indicus* (Ac) contig 9720 and *Drosophila melanogaster* (Dm) contig 996 were annotated as Iqf, a gene involved in many cellular functions including Notch signalling that is used in homeostasis and development (Csikós *et al.*, 2009; Overstreet *et al.*, 2003; Windler & Bilder, 2010). Ac contig 9783 and Dm contig 419 were annotated as spen, a gene involved in compound eye and nervous system development among other functions (Chang *et al.*, 2008; Mace & Tugores, 2004). Ac contig 19235 and Dm contig 18025 were annotated as Cg25C, a gene involved in producing an extracellular matrix structural constituent for cardiac muscle cell development and oviduct morphogenesis (FlyBase Curators *et al.*, 2004; Hollfelder *et al.*, 2014; Kelemen-Valkony *et al.*, 2012). Ac contig 12992 and Dm contig 5729 were annotated to CG14989 which has unknown functions. Ac contig 16700 and Dm contig 8900 were annotated to CG16700, a gene involved in amino acid transmembrane transport and neurogenesis (Romero-Calderon *et al.*, 2007). Ac contig 16152 and Dm contig 42643 were annotated to Hr4, a gene involved in development, growth and pupariation (King-Jones *et al.*, 2005).

Ac contig 18933 and Dm contig 4983 were annotated to tai, a gene involved in signal transduction and transcription regulation (Peyrefitte *et al.*, 2001). Ac contig 6926 and Dm contig 12959 were annotated to su(w[a]), a gene involved in mRNA splicing via spliceosome and RNA processing (FlyBase Curators *et al.*, 2004). Ac contig 12746 and Dm contig 525 were annotated to GA12325, a gene that produces a protein of unknown function with the domain DUF753 (FlyBase Curators *et al.*, 2004). Ac contig 15144 and Dm contig 17633 were annotated to Dyb, a gene involved in cytoskeletal protein binding and producing a muscle structural constituent (Goldstein & Gunawardena, 2000; Greener & Roberts, 2000). Ac contig 30904 and Dm contigs 1288 and 9885 were annotated to PMCA, a gene involved in calcium ion transmembrane transport and homeostasis (FlyBase Curators *et al.*, 2004; Roos *et al.*, 2005). Ac contig 11155 and Dm contig 24020 was annotated to mbl, a gene involved in the regulation of many biological processes and the development of the eye, muscle and nervous system (FlyBase Curators *et al.*, 2004).

The relative expression levels of each contigs were determined by the read count values listed in Table 2 (Fig. 43 & 44). In *Acletoxenus* cf. *indicus*, all the genes had moderate relative expression levels (Fig. 43). In *Drosophila melanogaster*, Iqf, spen, GA12352 and PMCA had very high relative expression levels. CG16700 had high relative expression levels and the remaining genes had moderate relative expression levels (Fig. 44). The protein domains that were mapped to the annotations were less than 50 amino acids long except for Cg25C and Dyb (Table 3). With the exception of PMCA, CG16700 and Mbl, the Ac contigs produced proteins that were very different from the Dm contigs as indicated by the pairwise distance that was more than 11% (Table 3).

Table 3. Summary of orthologous contigs between *Acletoxenus* cf. *indicus* and *Drosophila melanogaster*.

Gene	Acletoxenus cf. indicus Contig No. (Expression level)	*D. melanogaster* Contig No. (Expression level)	Aligned AA length	AA pairwise distance	Possible role
Plasma membrane Ca ATPase (PMCA)	Contig 30904 (Moderate expression)	Contig 1288 & Contig 9885 (Very high expression)	8	0.00%	Calcium ion transmembrane transport & homeostasis
Liquid facets (lqf)	Contig 9720 (Moderate expression)	Contig 996 (Very high expression)	19	25.10%	Cellular functions: Autophagy, endocytosis, neurotransmitter secretions, regulation of Notch signalling pathway, protein storage import into fat
Split ends (spen)	Contig 9783 (Moderate expression)	Contig 419 (Very high expression)	12	18.20%	Compound eye, peripheral nervous system development, axon guidance, positive regulation of wnt signalling pathway and other cellular functions
GA12325	Contig 12746 (Moderate expression)	Contig 525 (Very high expression)	16	13.40%	Domain of unknown function DUF753
CG16700	Contig 19145 (Moderate expression)	Contig 8900 (High expression)	12	0.00%	Amino acid transmembrane transport; neurogenesis
Collagen type IV (Cg25C)	Contig 19235 (Moderate expression)	Contig 18025 (Moderate expression)	137	33.60%	Extracellular matrix structural constituent, cardiac muscle cell development, oviduct morphogenesis
CG14989	Contig 12992 (Moderate expression)	Contig 5729 (Moderate expression)	18	19.40%	Unknown
Hr4	Contig 16152 (Moderate expression)	Contig 42643 (Moderate expression)	11	78.80%	Development, growth, pupariation, intracellular steroid hormone receptor signalling pathway, cellular response to hypoxia
Taiman (tai)	Contig 18933 (Moderate expression)	Contig 4983 (Moderate expression)	31	31.00%	Signal transduction, transcription regulation, axon extension, germ-line stem-cell niche homeostasis, border follicle cell migration
Suppressor of white-apricot (su(w[a]))	Contig 6926 (Moderate expression)	Contig 12959 (Moderate expression)	18	11.80%	mRNA splicing via spliceosome; RNA processing
Dystrobrevin-like (Dyb)	Contig 15144 (Moderate expression)	Contig 17633 (Moderate expression)	50	27.40%	Structural constituent of muscle, cytoskeletal protein binding, zinc ion binding.
Mbl	Contig 11155 (Moderate expression)	Contig 24020 (Moderate expression)	13	0.00%	Regulation of apoptosis, gene expression, female receptivity; eye, muscle and nervous system development, muscle cell homeostasis and other cellular functions

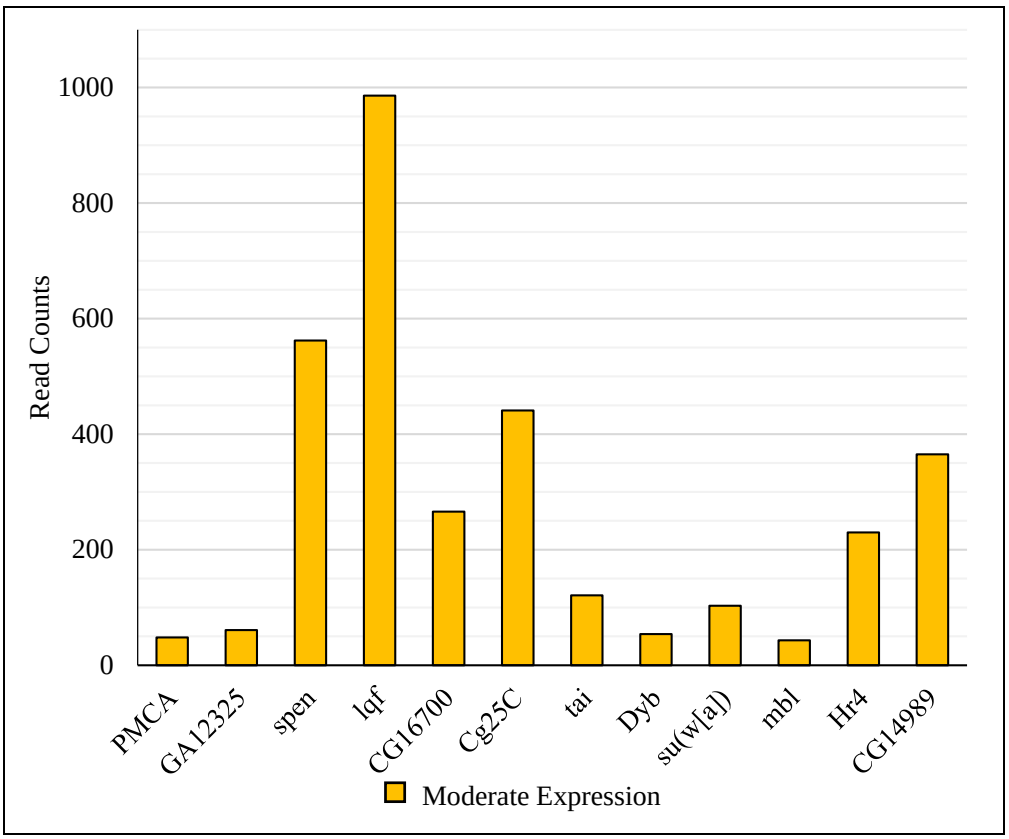

Fig. 43. Expression level of annotated contigs of *Acletoxenus* cf. *indicus*

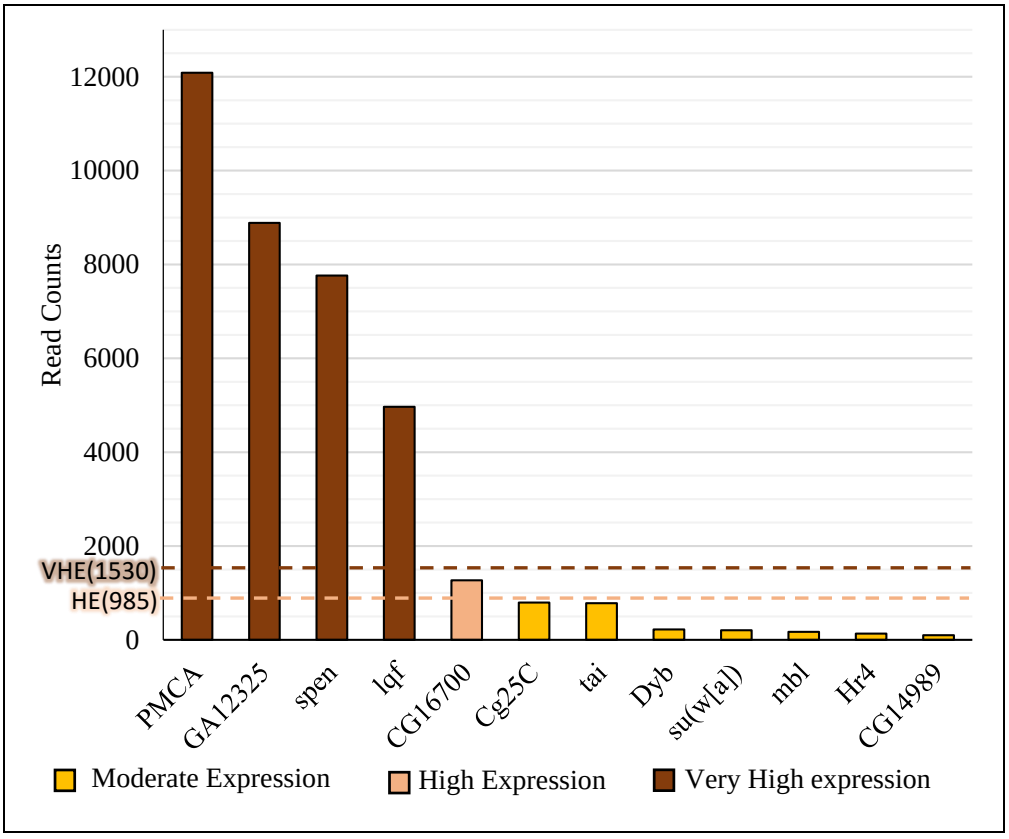

Fig. 44. Expression level of annotated contigs of *Drosophila melanogaster*.

3.4 Discussion

3.4.1 Comparison of Relatively Highly Expressed Contigs in *Acletoxenus* cf. *indicus* to The Drosophilidae Genome

Most of the contigs matched to genes with housekeeping and metabolic functions, but four contigs (108, 506, 16 and 2887) may have interesting functions related to the different life histories of *Drosophila melanogaster* and *Acletoxenus* cf. *indicus*. Contig 108 matched to the *Drosophila melanogaster* gene, Muc26B, which is only expressed in the egg of *Drosophila melanogaster* (Fisher *et al.*, 2012). The alignment length was 89 aligned amino acid residues with a pairwise distance of 57.5%. This indicated that the protein was quite different in the two species. As Muc26B is responsible for chitin metabolism in the egg stage, it could be possible that *Acletoxenus* cf. *indicus* has a modified version of Muc26B that is used in chitin catabolism to help in its digestion of its whitefly prey instead. However, the current evidence may be insufficient to prove that *Acletoxenus* sp. is able to digest the whole whitefly prey as none of the other contigs matched to any other hydrolytic enzymes.

Whitefly predators have been known to locate their prey through whitefly-induced plant volatiles (McGregor & Gillespie, 2004; Nomikou *et al.*, 2005). Thus, the high expression of a gene that might have odorant binding properties in contig 506 is interesting and not surprising. With a pairwise distance of 45.02%, the odorant that binds to expression of the Obp56d gene by *Drosophila virilis* is very different from the odorant being bound to the expression of contig 506 in *Acletoxenus* cf. *indicus*. This is not very surprising as *Drosophila virilis*

feeds on slime flux and decaying bark of tree (Seto & Tamura, 2013). As a result, both species would have their own set of odorant cues to locate their food.

Contig 16 potentially functioned to create a glue that allows larvae to attach themselves to a substrate during pupation. High expression of this gene occurs when the larva prepares to pupate. With a pairwise distance of 87.50%, the glue produced by *Drosophila virilis* is very different from that of *Acletoxenus* cf. *indicus* This is to be expected as the glue needed to attach to a wetter environment of decaying bark in *Drosophila virilis* during pupation would have to be very different from a generally dry environment at the abaxial surface of a leaf where *Acletoxenus* cf. *indicus* pupates (Spencer, 1938).

The function of the GJ24617 gene in *Drosophila virilis* is alkaloid biosynthesis, which could be similar for the orthologous contig 2887. Alkaloids can be used by insects as pheromones to attract a mate or for defence by being unpalatable or poisonous (Levinson, 1976; Schulz, 1998). Since the contig was expressed in the larva stage, its function is most likely for defence. However, this line of defence is not very effective as *Acletoxenus* cf. *indicus* was found to be heavily parasitized by *Pachyneuron leucopiscida* (Fig. 38).

On the other hand, since protein domain lengths are generally between 50 to 100 amino acid residues, it could be possible that contig 16 and 506 may have a different novel function from what was proposed as the alignment length of the amino acid residues were less than 50 (Table 2) (Xu & Nussinov, 1998). In addition, it could also be possible that the similarities and differences between the contigs and the Drosophila melanogaster genes were a result of neutral variation due to genetic drift and thus may not have the same function at all. On

the other hand, proteins with different primary structures can have the same function if they are regulated allosterically (Hervé, 1989). Thus, additional studies to study the motifs of the contigs would be needed to ascertain their functions.

3.4.2 Comparison of *Acletoxenus* cf. *indicus* and *Drosophila melanogaster* Larva Transcriptomes

The 12 *Acletoxenus* cf. *indicus* contigs that were confidently marked as orthologous to a *Drosophila melanogaster* gene showed only moderate levels of expressions in contrast to *Drosophila melanogaster* contigs, where Iqf, spen, GA12352 and PMCA had very high relative expression levels while CG16700 had high relative expression levels (Table 3; Fig. 44). It was not surprising that the genes Iqf (through Notch signalling) and PMCA which are involved in homeostasis would be highly expressed in *Drosophila melanogaster* but not in *Acletoxenus* cf. *indicus*. This is because *Drosophila melanogaster* larvae are found in harsh acidic and anaerobic environments as compared to the aerobic and dry environment where *Acletoxenus* cf. *indicus* is found. The higher expressions of the developmental genes Iqf (through Notch signalling), spen (compound eye and peripheral nervous system development) and CG16700 (neurogenesis) could be a result of the shorter life cycle of *Drosophila melanogaster*. As *Drosophila melanogaster* spends only 79 hours as a larva, one would very likely collect a larva that is expressing higher levels of developmental genes as compared to an *Acletoxenus* cf. *indicus* larva which has 12 days to develop (Fig. 26; Ashburner & Hawley, 2005)

However, the proposed functions for these contigs remain uncertain because the alignment length between the blast hit and *Acletoxenus* cf. *indicus* contigs were less than 50 except for the contigs that were identified as Dyb and Cg25C. In addition, the 27.4% and 33.6% pairwise distance of Dyb and Cg25C respectively indicated that close to a third of the amino acid residues were different and thus may have a different function from what is suggested by the blast hit.

3.4.3 Future studies

Although only 38 contigs were highlighted in this study, there were another 6028 *Acletoxenus* cf. *indicus* contigs that were identified as being so highly divergent from all other Drosophilidae gene entries on GenBank that they did not yield any hits. These contigs represented genes that have evolved so much, that the commonly used criteria for finding orthologues did not work (blastn $1e^{-5}$) (Li *et al.*, 2003). These 6028 contigs would be particularly good candidates for future studies. Unfortunately, a fully sequenced Steganinae genome would be required for one to determine their functions as the Drosophilinae genome may be too different. This again highlights the need to have more studies on Steganinae as there is currently very little information on this "other" subfamily of Drosophilidae (Otranto *et al.*, 2008)

Chapter 4. Drosophilidae Phylogeny

4.1 Introduction & Literature Review

Understanding the phylogenetic relationships of a group is essential in evolutionary biology and a pre-requisite for all comparative studies including embryology and speciation. Although *Drosophila* have been heavily used in such studies, there has been no firm consensus on the phylogenetic relationships of Drosophilidae and many key relationships remain controversial (Ashburner & Hawley, 2005; Markow & O'Grady, 2006). The first family-level phylogenetic analysis of Drosophilidae was by Throckmortan (1962, 1972) who divided the family into multiple radiations. However, his phylogenies contained few genera, predates the concept of strict monophyly and were not based on any explicit cladistics methods (Markow & O'Grady, 2006; Throckmorton, 1962; Throckmorton, 1975).

Okada (1989) analysed the phylogeny of Drosophilidae based on 14 morphological characters a few years later, where he proposed the subdivision of Drosophilidae into the subfamily Steganinae, consisting of the tribes Steganini and Leucophengini, and the subfamily Drosophilinae consisting of the tribes Microdrosophilini, Hypselethyrini, Colocasiomyini, Dettopsmyini and Drosophilini (Okada, 1989). This classification was revised by Grimaldi (1990) based on a cladistics analysis of 217 morphological characters where he found genus *Drosophila* to be paraphyletic (Fig. 45). He proposed to resolve this paraphyly by changing some ranks (Grimaldi, 1990). However, his placement of the endemic Hawaiian *Drosophila* species group remained very

controversial and was suggested to be a result of computational constraints (Remsen & O'Grady, 2002).

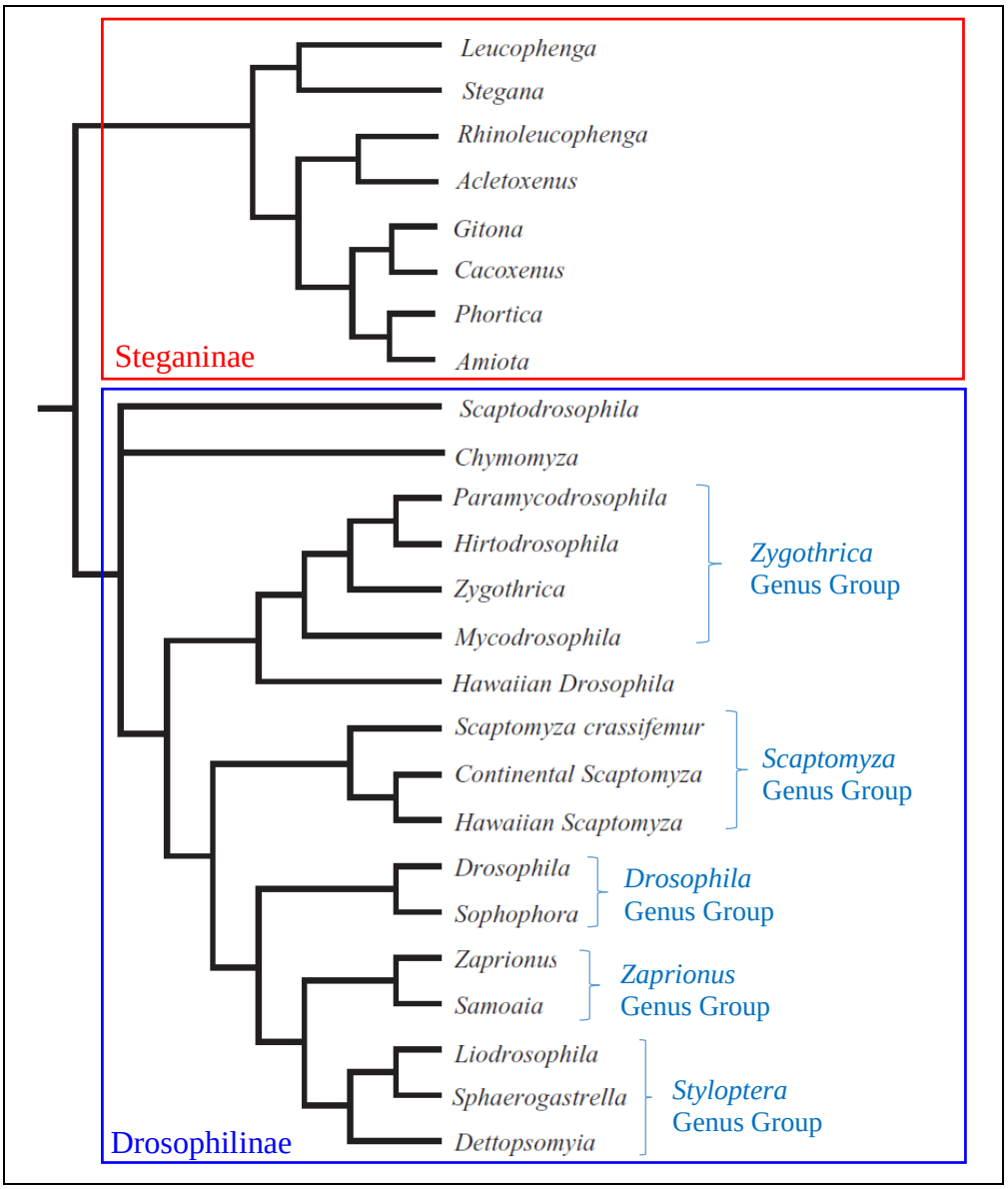

Fig. 45. Drosophilidae phylogeny based on Grimaldi (1990), modified from Markow & O'Grady (2006).

The following years saw phylogenies built from single gene trees, including 16S rRNA, (DeSalle, 1992) 28SrRNA (Pélandakis & Solignac, 1993), Adh (Russo *et al.*, 1995), SOD (Kwiatowski *et al.*, 1994) and Gpdh (Kwiatowski *et al.*, 1997). This was followed by phylogenies built from multiple genes (Gao *et al.*, 2011; O'Grady & DeSalle, 2008; Remsen & O'Grady, 2002; Tatarenkov *et al.*,

2001; Yassin *et al.*, 2010). Although some of these analyses yield robust support, these trees contained species mainly from the tribe Drosophilini and had topologies that were different. This lack of consensus could be a result of a lack of overlap in taxon sampling and gene coverage (van der Linde *et al.*, 2010). The two latest studies that sampled a larger number of taxa and genes was by Yassin (2013) who used 8 genes and morphological data for 330 taxa (Fig. 46 & 48) and van der Linde *et al*. (2010) who used 13 genes for 176 drosophilids (Fig. 48). As van der Linde *et al*. (2010) sampled very few Steganinae in her study, a summary of the Steganinae molecular phylogeny based on COI by Otranto (2008), was presented against Yassin's (2013) result in Fig. 46a. In addition, two morphological phylogenies of Steganinae by Chen (2000) and Grimaldi (1990) are presented in Fig. 47. A comparison of the Drosophilinae phylogeny between van der Linde *et al*. (2010) and Yassin (2013) is summarized in Fig. 48. As is clear from Fig. 45 to 48, only some relationships are congruent (especially in Drosophilinae), while others, especially in Steganinae are not. Thus, more studies are needed.

Genbank contains much data on Drosophilidae: There were 936,552 Drosophilidae nucleotide sequences available on the National Center for Biotechnology Information (NCBI) GenBank database as of 18 July 2015 (Benson *et al.*, 2013). This large amount of data could potentially resolve the Drosophilidae phylogeny by providing more gene and taxa coverage for the analysis. However, no one has yet attempted to make full use of the data despite the importance of *Drosophila* as models.

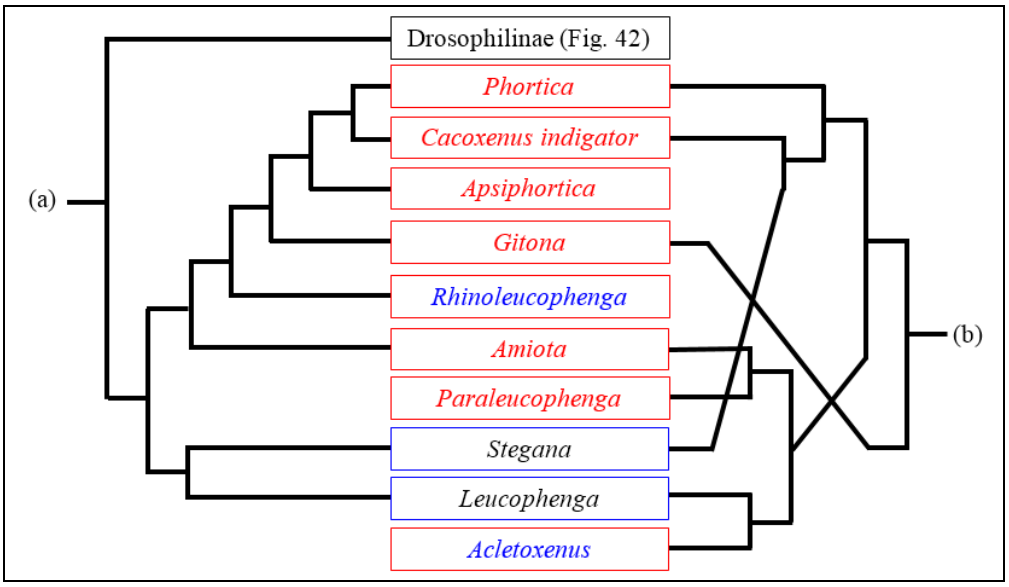

Fig. 46. Steganinae phylogeny based on (a) Yassin (2013) and (b) Otranto et al. (2008). Blue box = Steganini; Red box = Gitonini; Blue font = Acletoxenina; Red font = Gitonina.

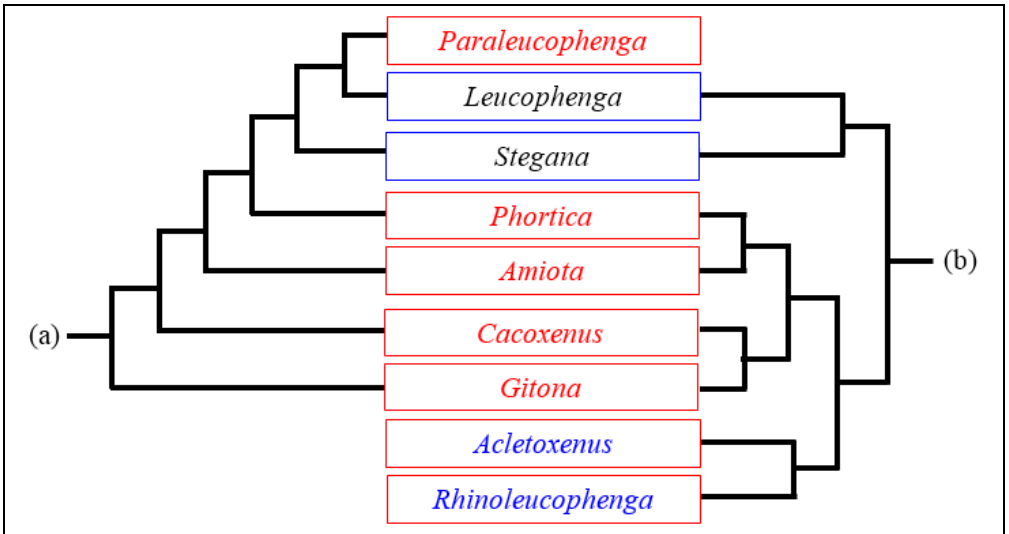

Fig. 47. Steganinae phylogeny based on (a) Chen (2000) and (b) Grimaldi (1990). Blue box = Steganini; Red box = Gitonini; Blue font = Acletoxenina; Red font = Gitonina.

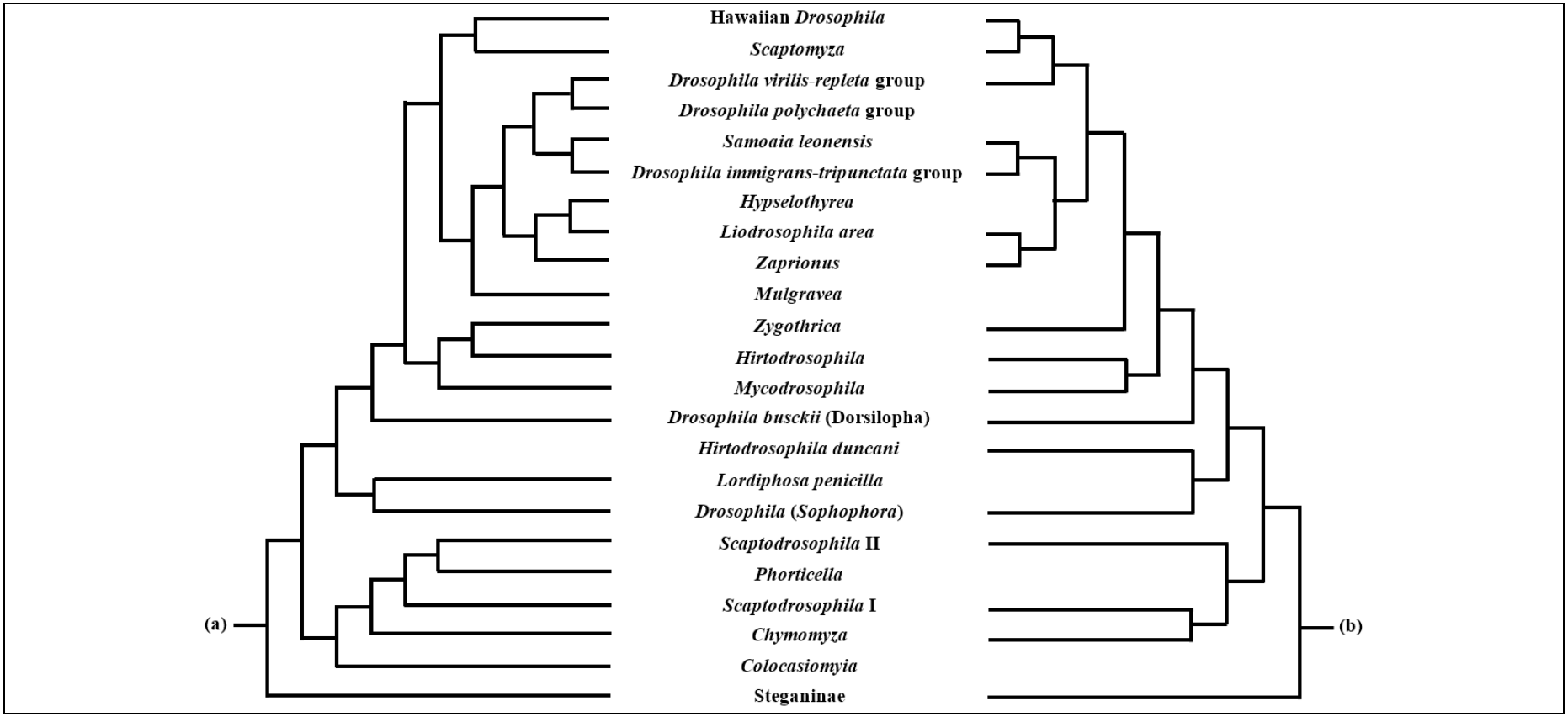

Fig. 48. Drosophilidae phylogeny based on (a) Yassin (2013) and (b) van der Linde *et al*. (2010). *Scaptodrosophila* I consist of *S. lebanonensis* and *S. stonei* in both papers in addition to *S. galloi* and *S. deflexa* in van der Linde *et al.* (2010). *Scaptodrosophila* II in Yassin (2013) consist of unidentified *Scaptodrosophila* species while *Scaptodrosophila* II from van der Linde *et al.* (2010) consist of. *S. dorsocentralis* and *S. latifasciaeformis*

Two methods frequently used to construct matrices based on large amount of data from different studies are the supertree and supermatrix methods. The supertree method consists of first conducting phylogenetic analysis on individual genes separately. The topologies of each gene tree are then converted into a binary matrix and combined together through algorithms such as matrix representation with parsimony (MRP). The combined matrix is then used to reconstruct the supertree (Sanderson *et al.*, 1998). Although supertrees have been criticized because they do not rely on the primary data, this method is less computationally expensive and has been able to generate highly accurate estimations of species trees (Bininda-Emonds, 2005; Yang & Warnow, 2011). In contrast, the supermatrix method concatenates all the genes into one matrix for a phylogenetic analysis (de Queiroz & Gatesy, 2007). Thus, the major advantage of the supermatrix method over the supertree method is that the primary data is more fully utilized in tree building (de Queiroz & Gatesy, 2007). However, supermatrices usually contain a huge amount of missing characters and analysis is more computationally expensive (de Queiroz & Gatesy, 2007; McMahon & Sanderson, 2006).

After forming a matrix, one would have to choose an optimality criterion for selecting trees. Two commonly used criteria are maximum parsimony and maximum likelihood. Maximum parsimony assigns character states to interior nodes on a tree to minimize the number of changes on a phylogenetic tree (Yang & Rannala, 2012). Maximum likelihood finds the model consisting of tree, branch lengths, nucleotide frequencies and substitution rates that explains the data with the highest likelihood (Felsenstein, 1981). As both criteria contain their own set of advantages and disadvantages (Bininda-Emonds, 2005; de

Queiroz & Gatesy, 2007; McMahon & Sanderson, 2006; Yang & Warnow, 2011), both criteria are used here for tree building.

Finally, low quality alignments have been shown to have a huge impact on the final phylogenetic trees (Morrison & Ellis, 1997; Ogden & Rosenberg, 2006; Smythe *et al.*, 2006). Thus, alignment masking software, such as Gblocks and Zorro, is here used to test whether resolution and support are a function of alignment quality.

In this chapter, I aimed to:

1. Resolve the Drosophilidae phylogeny using data from GenBank to increase gene and taxa coverage.
2. Compare the results produced by the supermatrix and supertree approaches.
3. Compare the results produced with and without alignment masking
4. Place *Acletoxenus* onto the tree-of-life for Drosophilidae

4.2 Material and Methods

4.2.1 Extraction of Genes

The data used in the analyses were downloaded from NCBI GenBank in the Nucleotide database using the search term "txid7214[Organism] OR txid27457[Organism] OR txid7213[Organism] OR txid7394[Organism] OR txid141453[Organism] OR txid139649[Organism] OR txid141453[Organism] OR txid292399[Organism]" on 7 April 2014. This downloaded 1,261,085 nucleotide entries from GenBank filed under Drosophilidae in addition to the outgroups of *Bactrocera dorsalis* (Tephritidae), *Ceratitis capitata* (Tephritidae), *Glossina morsitans* (Glossinidae), *Sepsis cynipsea* (Sepsidae), *Teleopsis dalmanni* (Diopsidae), *Teleopsis whitei* (Diopsidae) and *Themira biloba* (Sepsidae).

A custom Node.js® script was then used to pull out the full list of unique gene and product names and the number of species present to each name. From this list of 2710 names, only those with more than 50 species were retained resulting in a list of 31 genes: COI, COII, COIII, 12S rRNA, 16S rRNA, 28SrRNA, ADH, Amyrel, per, Yp1, Ddc, Amy, Adhr, amd, ATP6, ATP8, Cyt-b, esc, fkh, Gpdh, H2a, Hb, marf, ND2, ND3, ND4, ND5, PTC, RpL32, snf, wee. A filter was then created for each gene containing its synonyms, annotations and product names. The filters containing the synonym for each gene is found in Appendix, Table A5 (pg. A-8 to A-9). Another custom Biopython script was used to look for these terms in the "Gene name", "Product name" and "Note" headers in the downloaded file to extract all the sequences for a specific gene. The results were

fasta files containing all entries for each gene, which will be referred to as "gene fasta files" subsequently.

As there were problems in aligning 28S rRNA genes in the later steps since the gene entries were too fragmented, an additional step to extract smaller regions of this gene was conducted. Looking under the "Notes" header for each entry in the downloaded file, sequences that were annotated as the domains of 28Sd1, 28Sd2, 28Sd3 and 28Sd10 were extracted with a custom Node.js® script in fasta format. These sequences listed in Appendix, Table A6 (pg. A-10) were then used to query for each domain against a database containing 28S rRNA genes using the BLAST algorithm (minimum similarity of $1e^{-5}$ and word size 9) aligning with MAFFT ver. 7 (Altschul *et al.*, 1990; Bocak *et al.*, 2014; Katoh & Standley, 2013). The result was a fasta file containing entries that were homologous for each domain of 28S rRNA, which will be referred to as "28S rRNA domain fasta files".

In order to obtain the 34 gene loci for *Acletoxenus* cf. *indicus*, the contigs from Section 3.2.1 was blasted to the database of the 34 gene loci using the blastn algorithm with minimum similarity of $1e^{-5}$ (Altschul *et al.*, 1990). The contigs that matched the gene loci were then aligned to the respective gene of *Drosophila melanogaster* and concatenated together with "?" in order to fill in the missing regions. Thereafter, they were added to their respective gene and 28S rRNA domain fasta files.

4.2.2 Alignment of Genes

The protein-encoding gene fasta files were aligned with using Translator X using MAFFT ver. 7 on default settings (Abascal *et al.*, 2010; Katoh & Standley,

2013). 12S rRNA and 16S rRNA gene fasta files were aligned using MAFFT ver. 7 using default settings while 28S rRNA domain fasta files were aligned using MAFFT ver. 7 with Q-INS-I settings (opening cost of 1.9 and extension cost 0.2). As the period (per) gene had alignments that were not well aligned even with different MAFFT alignment settings (G-INS-I, E-INS-I, FFT-NS-i), it was removed from subsequent analyses. Subsequent analyses are thus based on 33 gene loci: COI, COII, COIII, 12S rRNA, 16S rRNA, 28Sd1, 28Sd2, 28Sd3, 28Sd10,ADH, Amyrel, Yp1, Ddc, Amy, Adhr, amd, ATP6, ATP8, Cyt-b, esc, fkh, Gpdh, H2a, Hb, marf, ND2, ND3, ND4, ND5, PTC, RpL32, snf, wee.

4.2.3 Unique Gene Fasta File and Quality Check

From the alignments of each gene fasta file, the longest sequence was chosen to represent a species if multiple GenBank entries were found for the same gene region. If there were multiple entries found in different regions in the alignments for one species, they were concatenated with the missing data coded as "?". This created an aligned fasta file with one sequence entry to represent one species for a specific gene, which will be referred to as unique gene fasta file.

Gene trees used for quality checks were created by first creating a TNT sequence matrix with SequenceMatrix ver. 1.7.8 for each aligned gene fasta file. Parsimony trees were built from these sequence matrix files based on scripts by Simmons & Goloboff (2014) for high quality searches that built one hundred trees before hybridizing them in TNT ver. 1.1 (Simmons & Goloboff, 2014). One hundred bootstrap replicates based on scripts by Simmons & Goloboff (2014) for relatively thorough searches were then performed and mapped onto

the strict consensus of the most parsimonious trees, from the tree hybridization stage, using SumTrees 3.3.1 of the DendroPy ver. 3.8.0 package.

As GenBank may include sequences from misidentified specimens, quality checks were performed on the unique gene fasta files for protein encoding genes through (1) pairwise distance checks and (2) looking at the relationships in the gene trees of the aligned gene fasta files. First, inter-specific pairwise distance that were less than 1% were identified from the unique gene fasta files using TaxonDNA/Species Identifier ver. 1.7.7 (Meier *et al.*, 2006). These pairs were then located on the gene tree in order to determine position on the tree and bootstrap support. Any pair that had a pairwise distance of 0% with a bootstrap support of 100% were deemed to be questionable as different species should not have fully identical sequences for fast-evolving genes.

Very similar sequences with low pairwise distance in different species are expected for slowly evolving genes. In these cases, the support values for these species in a gene tree would be very low. Thus, any pair that had a genetic distance of more than 0% but less than 1% with bootstrap values more than 85% were deemed as questionable unless (1) they could not interbreed; (2) the same pair was not present in the other gene loci; (3) the same pair was present in previous published phylogeny analyses. The 1% criteria for pairwise distance was chosen as it is a common criteria used to delimit closely related species (Brower, 1994; Schloss & Handelsman, 2006).

In those cases where sequence identities were in doubt, the species were checked for whether they had a different sequence entry that would result in the pairwise distance > 1%. If there were, the other sequence entry was picked to

represent the species. For those species that lacked such an entry, the number of sequences of a gene for each of the two species were determined. If there was a different number of sequence entries for the same gene, the species that had a higher number of sequence entries was kept as it was more likely to be correct. On the other hand, if both species had the same number of sequence entries, both were deleted from the dataset as one would not be able to tell which species was misidentified.

rRNA genes were excluded from this check as there were thousands of pairs that had pairwise distance < 1% and the gene trees were collapsed and unresolved. This indicated that the low inter-specific differences in the rRNA genes were not likely due to contamination.

4.2.4 Supertree (Maximum Likelihood) Approach

The quality checked gene fasta files that were mitochondrial (COI, COII, COIII, ATP6, ATP8, Cytb, ND2, ND3, ND4, ND5) were concatenated with SequenceMatrix ver. 1.7.8 and exported in newick format (Vaidya *et al.*, 2011). The 4 domain files of 28S rRNA were also concatenated and exported. The remaining gene fasta files were also converted to newick format. The newick files were then uploaded to the Cyberinfrastructure for Phylogenetic Research (CIPRES) website (Miller *et al.*, 2010). Using CIPRES, a gene tree was built based on each gene newick file using RAxML on XSEDE, using the GTRCAT model (Stamatakis, 2014). Thereafter, the gene trees were opened in Mesquite ver. 2.73 to build the matrix representation parsimony matrices (MRP) (Maddison & Maddison, 2015).

Some taxa can be phylogenetically unstable due to missing or limited data. These rogue taxa often have negative impact on topological resolution and support (Sanderson *et al.*, 1998; Wilkinson, 1995, 2003). Thus, safe taxonomic reduction (Wilkinson, 1995) was used as implemented in the Concatabomination pipeline (Siu-Ting *et al.*, 2015) which uses the MRP to determine rogue taxa. After removing rogue taxa from the MRP, RAxML on XSEDE, using the GTRCAT model was used for tree building and calculating branch-support values based on 100 bootstraps.

An alternate approach to identify rogue taxa was also carried out by first generating the best tree and 100 bootstrap trees with RAxML on XSEDE, using the GTRCAT model of the MRP on CIPRES (Miller *et al.*, 2010; Stamatakis, 2014). The best tree and 100 bootstrap trees were then used to search for rogue taxa using the RogueNaRok algorithm with majority-rule consensus as the threshold with a maximum drop set size of 1. After removal of the rogue taxa generated by RougeNaRok from the MRP, RAxML on XSEDE, using the GTRCAT model was then used for building the tree and generating 100 bootstrap on CIPRES (Miller *et al.*, 2010; Stamatakis, 2014). Another rogue taxa identification program that was tried and abandoned was IterPCR. This program was unable to produce any results even after 2 months of computing and was terminated as it was taking too long.

4.2.5 Supertree (Maximum Parsimony) Approach

Each quality checked gene fasta file was converted to TNT format with SequenceMatrix ver. 1.7.8 (Vaidya *et al.*, 2011). The consensus of the most parsimonious trees for each gene were built with the same scripts mentioned in

Section 4.2.3 with TNT ver. 1.1 (Goloboff *et al.*, 2008; Simmons & Goloboff, 2014). These gene trees were then combined and uploaded into Mesquite ver. 2.73 to build the matrix representation parsimony (MRP) (Maddison & Maddison, 2015). A list of rogue taxa was then generated using the Concatabomination pipeline (Siu-Ting *et al.*, 2015). After removing the rogue taxa from the MRP, tree building and 100 bootstraps were generated using the same scripts mentioned in Section 4.2.3 with TNT ver. 1.1 (Goloboff *et al.*, 2008; Simmons & Goloboff, 2014). The bootstraps values were then mapped onto the strict consensus of the most parsimonious trees using SumTrees ver. 3.3.1 of the DendroPy ver. 3.8.0 package (Sukumaran & Holder, 2010).

As RogueNaRok only runs on fully resolved bootstrap trees, it could not be used on the maximum parsimony bootstrap trees which may be collapsed at some nodes. Thus, the rogue taxa generated from RogueNaRok from section 4.2.4 were removed from the MRP under the assumption that the rogue taxa would be similar in the ML and MP analyses. After removal of the rogue taxa, tree building and 100 bootstraps were generated using the same scripts mentioned in Section 4.2.3 with TNT ver. 1.1 (Goloboff *et al.*, 2008; Simmons & Goloboff, 2014). The bootstraps values were then mapped onto the strict consensus of the most parsimonious trees using SumTrees ver. 3.3.1 of the DendroPy ver. 3.8.0 package (Sukumaran & Holder, 2010).

4.2.6 Supermatrix (Maximum Likelihood) Approach

The quality checked gene fasta files were concatenated and the matrix exported in newick format using SequenceMatrix ver. 1.7.8 (Vaidya *et al.*, 2011). Attempts to generate a list of rogue taxa from the concatenated supermatrix with

the Concatabomination pipeline (Siu-Ting *et al.*, 2015) and IterPCR (Pol & Escapa, 2009) were abandoned as the processes could not handle the big dataset and were still running after a month. Thus, RogueNaRok was used to generate a list of rogue taxa (Aberer *et al.*, 2011). The RogueNaRok method consisted of first building a best known tree and 200 bootstrap trees using RAxML on XSEDE, using the GTRCAT model from the supermatrix. The best tree and bootstraps were then used to search for rogue taxon using the RogueNaRok algorithm with majority-rule consensus as the threshold with a maximum drop set size of 1. After removal of the rogue taxa, RAxML on XSEDE, using the GTRCAT model was used for building the best tree and generating 200 bootstrap on CIPRES (Miller *et al.*, 2010; Stamatakis, 2014).

4.2.7 Supermatrix (Maximum Parsimony) Approach

The quality checked fasta files were concatenated and the matrix exported in TNT format with SequenceMatrix ver. 1.7.8 (Vaidya *et al.*, 2011). Similar to Section 4.2.6, a list of rogue taxa could not be generated from the concatenated supermatrix with the Concatabomination pipeline (Siu-Ting *et al.*, 2015) and IterPCR (Pol & Escapa, 2009). As RogueNaRok only runs on fully resolved bootstrap trees, the rogue taxa generated from Section 4.2.6 were removed from the TNT matrix under the assumption that the rogue taxa would be similar in the ML and MP analyses. After removal, tree building and bootstrapping were conducted with the same scripts mentioned in Section 4.2.3 with TNT ver. 1.1, except that a thousand trees were replicated for tree hybridization during the tree building stage (Goloboff *et al.*, 2008; Simmons & Goloboff, 2014). The bootstraps values were then mapped onto the strict consensus of the most

parsimonious trees using SumTrees ver. 3.3.1 of the DendroPy ver. 3.8.0 package (Sukumaran & Holder, 2010).

4.2.8 Alignment Masking With Supermatrix

In an attempt to reduce the impact of poor-quality alignments, two commonly used algorithms, Gblocks and Zorro, were used to mask ambiguous alignments of the rRNA gene fasta files. Alignment masking was not carried out for the protein encoding genes as the resulting alignments would change translations due to frame shift and will not represent the actual condition. The default parameters were used for Gblocks except the length of flanking regions was set to 50% and all gaps were allowed (Talavera & Castresana, 2007). The cut-off used for Zorro to remove questionable nucleotide positions was set at 1.5 instead of the default of 4.0 (Wu *et al.*, 2012). The relaxed criteria for the alignment masking were chosen as the default resulted in largely unresolved trees during a test run that built a hundred maximum parsimony trees.

The two settings of masked files were then concatenated with the other quality checked gene fasta files using SequenceMatrix ver. 1.7.8 separately to create two supermatrices (Vaidya *et al.*, 2011). The two supermatrices were then used to build a maximum likelihood tree with 200 bootstraps using the same methods as Section 4.2.5 and a maximum parsimony tree using the same methods as Section 4.2.6. As the maximum parsimony supertree approach was giving many unresolved nodes in the trees, alignment masking was not carried out with the supertree approaches.

4.2.9 Post-Phylogenetic Analysis Rogue Taxa Check

There were a few individuals that did not nest with other individuals of the same classification found on www.taxodros.uzh.ch/ by Bächli (2015). These taxa were checked for whether they were rogue taxa using RogueNaRok and IterPCR. The same program could not be used on both type of trees to determine rogue taxa as RogueNaRok only works on fully bifurcated trees but the maximum parsimony trees would collapse at unsupported nodes. Furthermore, multiple most parsimonious trees were found by the maximum parsimony approach but only one best tree would be produced by the maximum likelihood approach. Thus, IterPCR could input the most parsimonious trees from the maximum parsimony analyses to locate for rogue taxa but only bootstraps could be input to locate for rouge taxa for the maximum likelihood analyses. As mentioned previously, IterPCR was ultimately not used for the maximum likelihood analyses because it was computationally too expensive. Therefore, IterPCR using the most parsimonious trees were used to determine rouge taxa for the maximum parsimony analyses while RogueNaRok using strict consensus threshold with a maximum drop set size of one was used to determine rouge taxa for the maximum likelihood analyses. Concatabomination could not be carried out as the supermatrices were too big for the program to run.

The interspecific pairwise distances between *Pseudostegana* species were calculated for the five gene loci (COI, ND2, 28Sd1, 28Sd2 & 28Sd3) that contained this genus as *Pseudostegana bilobata* was not identified as a rogue taxa during the post-phylogenetic analysis rogue taxa check. Although *Gitona pualiani* was also not identified as a rogue taxa in the post-phylogenetic analysis

rogue taxa check, the interspecific pairwise distances for *Gitona* could not be calculated as it was the only species from that genera in the analyses.

4.3 Results

4.3.1 Extraction, Alignment and Quality Check of Gene Files

A total of 6,363 sequences from 1,085 species were chosen for the phylogenetic analysis of Drosophilidae after quality checks. The NCBI ascension numbers for these sequences can be found in the online supplementary material, SM01, at https://goo.gl/soTHJk. Out of the 34 gene loci, 22 gene loci were matched to the assembled *Acletoxenus* cf. *indicus* contigs. The sequences are listed in Appendix, Table A7 (pg. A-11 to A-14). The final alignments of the 33 gene loci use in subsequent analysis excluding the period gene is found in the online supplementary material, SM02, at https://goo.gl/soTHJk.

4.3.2 Supertree Approach

The MRP generated from the mitochondria, 12SrRNA, 16SrRNA, 28S rRNA, ADH, Amyrel, Yp1, SOD, Ddc, Amy, Adhr, Amd, esc, fkh, Gpdh, H2a, marf, PTC, RpL32, snf and wee maximum likelihood (ML) trees contained 1,085 species with 3,879 columns of binary characters containing 62.6% missing characters (online supplementary material SM03 at https://goo.gl/soTHJk). The MRP generated from the maximum parsimony (MP) gene loci trees contained 1,085 species with 2,820 binary characters containing 61.4% missing characters (online supplementary material SM04 at https://goo.gl/soTHJk). The Concatabomination (Cat) pipeline found 253 species (including *Zygothrica* species, *Lordiphosa acutissima* and *Lordiphosa pseudotenuicauda*) and 156 species (including *Zygothrica* species and *Lordiphosa tenuicauda* species group) to be unstable in the MP and ML approaches respectively while the RogueNaRok (Rog) pipeline found 242 species that were unstable (online

supplementary material SM05 at https://goo.gl/soTHJk). Only 14 taxa were commonly identified as rogue taxa in both the RogueNaRok and Concatabomination programs while the Concatabomination program found 104 common rogue taxa between the MP and ML MRP (online supplementary material SM05 at https://goo.gl/soTHJk).

After removal of the rogue taxa by Concatabomination, the ML MRP contained 929 taxa with 3,879 characters containing 59.6% missing characters while the MP MRP contained 832 taxa with 2,820 characters containing 56.7% missing characters (online supplementary materials SM06 & SM07 at https://goo.gl/soTHJk). The ML MRP, after removal of rogue taxa by RogueNaRok, contained 843 taxa with 3,879 characters containing 61.6% missing characters while the MP MRP contained 843 taxa with 2,820 characters containing 60.3% missing characters (online supplementary materials SM08 & SM09 at https://goo.gl/soTHJk).

After removing the rogue taxa detected by Cat, the location where most of the individuals of the same genus clustered together was summarized in Fig. 49. The full MP and ML trees can be found in the online supplementary material SM10 and SM11 respectively at https://goo.gl/soTHJk. The summary after removal by Rog using MP and ML approaches was mapped in Fig. 50 and the full MP and ML trees can be found in the online supplementary material SM12 and SM13 respectively at https://goo.gl/soTHJk. Regardless of the methods to remove rogue taxa, the MP supertree approach was unable to resolve any relationships of Drosophilidae (Fig. 49 & 50). The Concatabomination ML tree showed Drosophilidae to be monophyletic while the RogueNaRok ML tree showed Drosophilidae to be paraphyletic as the outgroup *Teleopsis whitei* was

nested within Drosophilidae (Fig. 49 & 50). At the subfamily level, Steganinae was not monophyletic. In the Cat ML supertree, *Colocasiomyia* (Drosophilinae) was nested with *Rhinoleucophenga* and *Gitona pauliani*, which were classified under Steganinae (Fig. 49). In the Rog ML supertree, the monophyletic Drosophilinae had low support (bootstrap = 24%) and was nested within Steganinae (Fig. 50). The relationships of Steganinae were similar in both ML trees but did not correspond to the current classification of tribes Steganini (*Leucophenga* and *Stegana*) and Gitonini (other remaining groups of Steganinae) (Fig. 49 & 50). In both ML trees, the Drosophilinae Drosophilini subtribes, Colocasiomyina (*Colocasiomyia*) and Drosophilina (other remaining groups of Drosophilinae) were monophyletic with a bootstrap support values of 83% and 2% respectively in the Cat tree (Fig. 49) and 76% and 24% respectively in the Rog tree (Fig. 50). The following clades were monophyletic in both ML trees (Fig 49 & 50): *Phortica* (Cat bootstrap = 76%, Rog bootstrap = 68%), *Apsiphortica* (Cat bootstrap = 68%, Rog bootstrap = 93%), *Colocasiomyia* (Cat bootstrap = 83%, Rog bootstrap = 76%), *Drosophila polychaeta* species group (Cat bootstrap = 66%, Rog bootstrap = 44%), *Scaptomyza* (Cat bootstrap = 28%, Rog bootstrap = 91%) and the Hawaiian *Drosophila* species group (Cat bootstrap = 4%, Rog bootstrap = 78%). *Acletoxenus* cf. *indicus* was a sister clade to *Parastegana* and *Pseudostegana* in the Rog ML tree (Fig. 50) but a sister clade to the other Drosophilidae in the Cat ML tree (Fig. 49).

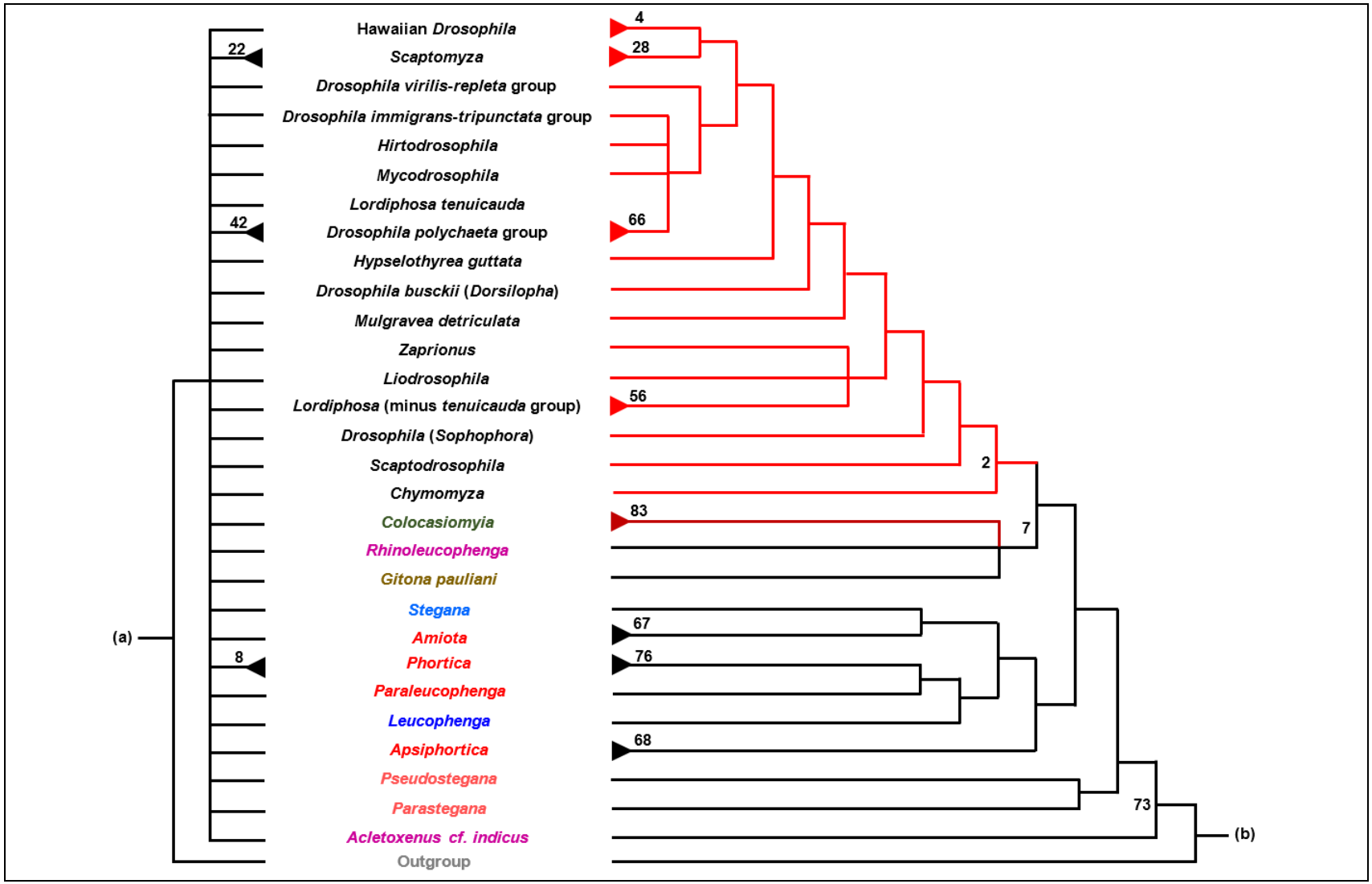

Fig. 49. Drosophilidae phylogeny using the Supertree approach after removal of rogue taxa by Concatabomination using (a) MP and (b) ML.

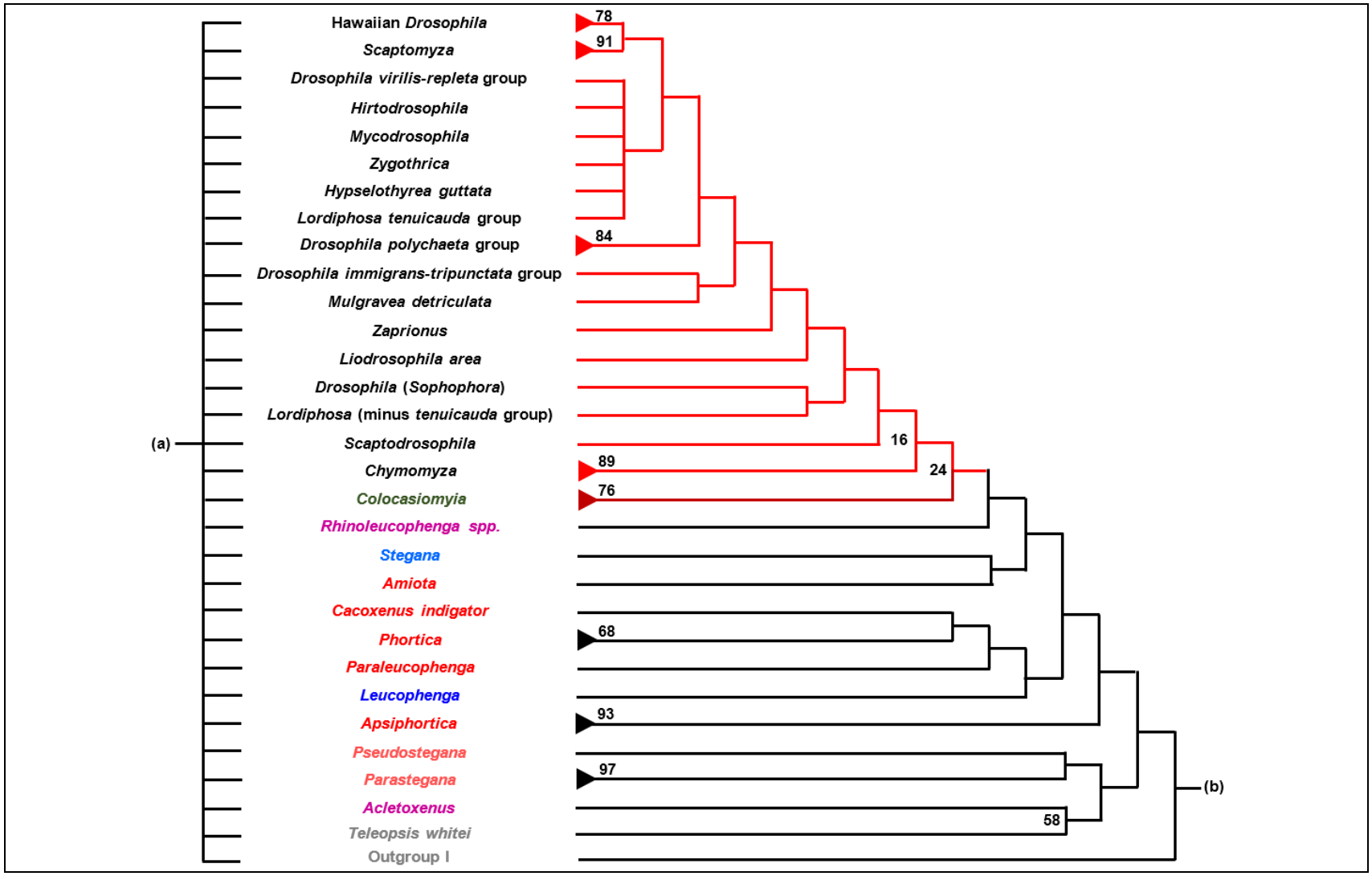

Fig. 50. Drosophilidae phylogeny using the Supertree approach after removal of rogue taxa by RogueNaRok using (a) MP and (b) ML.

4.3.3 Supermatrix Approach

The supermatrix created from the concatenation of COI, COII, COIII, ATP6, ATP8, Cytb, ND2, ND3, ND4, ND5, 12SrRNA, 16SrRNA, 28S rRNA, ADH, Amyrel, Yp1, SOD, Ddc, Amy, Adhr, Amd, esc, fkh, Gpdh, H2a, marf, PTC, RpL32, snf and wee genes contained 1,086 species with 31,437 columns of nucleotide characters containing 87.5% missing characters (online supplementary materials SM14 at https://goo.gl/soTHJk). The RogueNaRok pipeline found 207 species (including *Acletoxenus formosus* and *Leucophenga zhenfangae*) that were unstable (online supplementary materials SM15 at https://goo.gl/soTHJk). After removing the 207 rogue taxa, the supermatrix contained 879 species with 31,437 characters containing 86.3% missing characters (online supplementary materials SM16 at https://goo.gl/soTHJk). The resulting phylogenies by the MP and ML approaches were presented in Fig. 51 and 52 while the full trees can be found in the online supplementary materials SM17 and SM18 at https://goo.gl/soTHJk . Fig. 51 shows the relationships in the Steganinae radiation while Fig. 52 shows a summary of where the most individuals of a genera in the Drosophilinae radiation was found. Between the supermatrix and supertree approaches, there were only 64 rouge taxa generated by RogueNaRok that were common (online supplementary material SM15 at https://goo.gl/soTHJk).

Overall, there were a few conflicts in relationships generated between the MP and ML approaches (Fig. 51 & 52). Both approaches found Drosophilidae to be monophyletic (MP bootstrap = 32% & ML bootstrap = 82%; Fig. 51). At the subfamily level, Steganinae and Drosophilinae was paraphyletic in the MP analysis. *Colocasiomyia* (Drosophilinae) was nested with *Acletoxenus*

(Steganinae), although the bootstrap support values were quite low at 7% (Fig. 51). In the ML analysis, the monophyletic Drosophilinae (bootstrap = 59%) was nested within Steganinae with a low bootstrap value of 6% (Fig. 51).

Only one subtribe, Colocasiomyina (*Colocasiomyia*) was monophyletic (MP bootstrap = 40%, ML bootstrap = 59%; Fig. 51). The Steganinae Gitonini subtribe Gitonina (which includes *Amiota*, *Apsiphortica*, *Cacoxenus*, *Gitona*, *Paraleucophenga*, *Phortica*) were all clustered together with the exception of *Gitona pauliani* (Fig. 51) *Rhinoleucophenga* and *Acletoxenus*, which were classified under the Steganinae Gitonini subtribe, Acletoxenina, did not group together (Fig. 51). The relationships between the genera in Drosophilinae were different between the clades of (*Drosophila immigrans-tripunctata* species group + *Drosophila polychaeta* species group), (*Zaprionus* + *Liodrosophila* + *Hypselothyrea guttata*), *Lordiphosa tenuicauda* species group and (*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*) (Fig. 52). In the MP approach, *Lordiphosa tenuicauda* species *group* and (*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*) formed a sister clade to the clade of (*Zaprionus* + *Liodrosophila* + *Hypselothyrea guttata*). The clade consisting of the two sister clades were in turn a sister clade to (*Drosophila immigrans-tripunctata* species group + *Drosophila polychaeta* species group). The ML approach in contrast has (*Drosophila immigrans-tripunctata* species group + *Drosophila polychaeta* species group) as a sister clade to the clade of (*Zaprionus* + *Liodrosophila* + *Hypselothyrea guttata*). These two sister clades formed a clade that was a sister group to a clade of [*Lordiphosa tenuicauda* species group, (*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*)].

The species under each Steganinae genera were generally nested together in both trees with the exception of *Pseudostegana bilobata*, *Stegana mehadiae, Leucophenga abbreviata* species group (consisting of *Leucophenga brevivena*, *Leucophenga abbreviata* and *Leucophenga sujuanae*) and *Phortica picta* (Fig. 51). The following clades below the genera level were monophyletic in both trees (Fig 51 & 52): *Acletoxenus* (MP bootstrap = 85%, ML bootstrap = 100%), *Apsiphortica* (MP bootstrap = 63%, Rog bootstrap = 97%), *Parastegana* (MP bootstrap = 99%, Rog bootstrap = 100%), *Colocasiomyia* (MP bootstrap = 40%, ML bootstrap = 95%), *Drosophila polychaeta* species group (MP bootstrap = 88%, ML bootstrap = 100%), *Scaptomyza* (MP bootstrap = 28%, ML bootstrap = 91%) and the Hawaiian *Drosophila* species group (MP bootstrap = 4%, ML bootstrap = 78%).

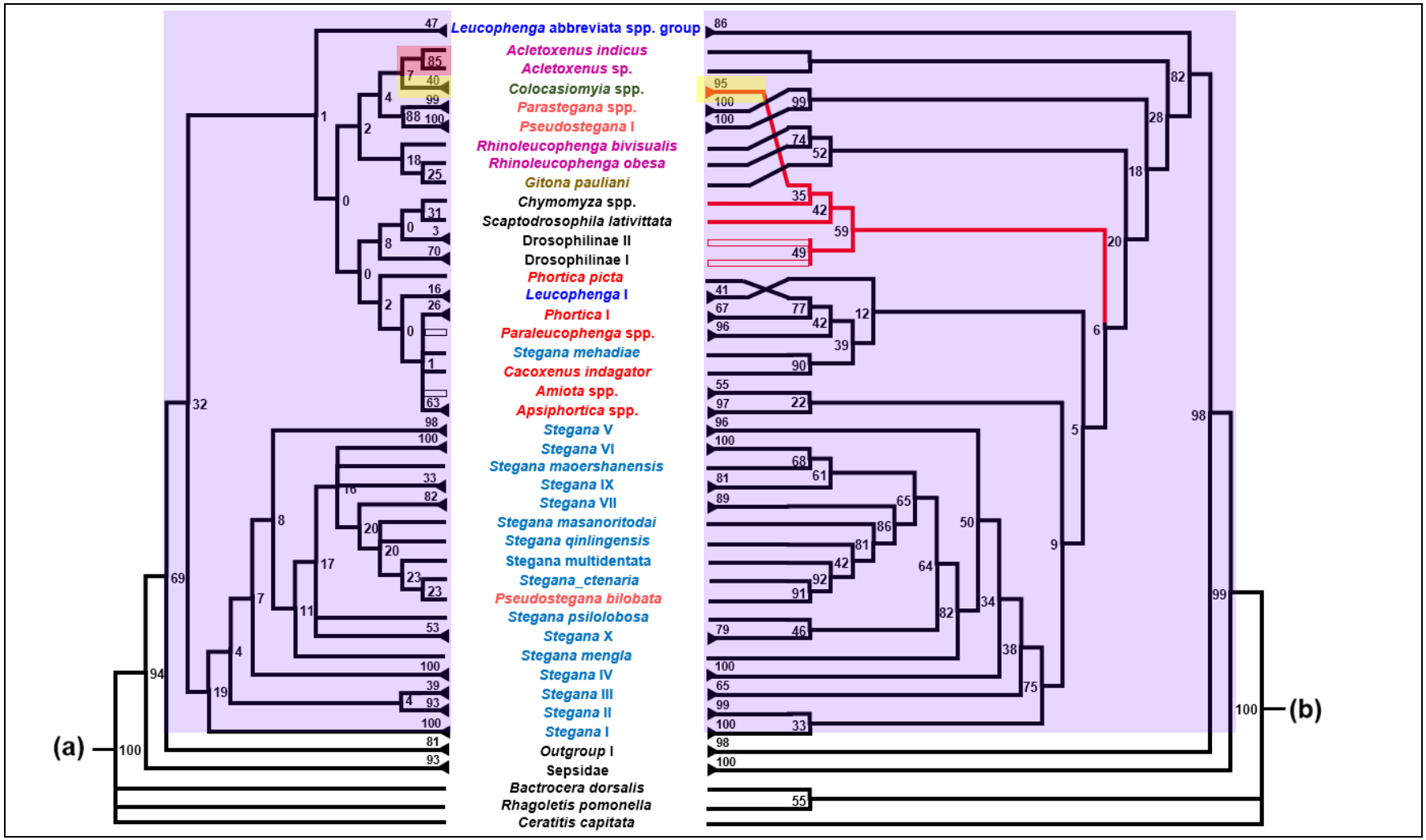

Fig. 51. Drosophilidae (Steganinae) phylogeny using the Supermatrix approach after removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. ▭ : Unresolved collapsed phylogeny, ▭ : Drosophilidae, ▭ : Drosophilinae, ▭ : Colocasiomyina, ▭ : *Acletoxenus.*

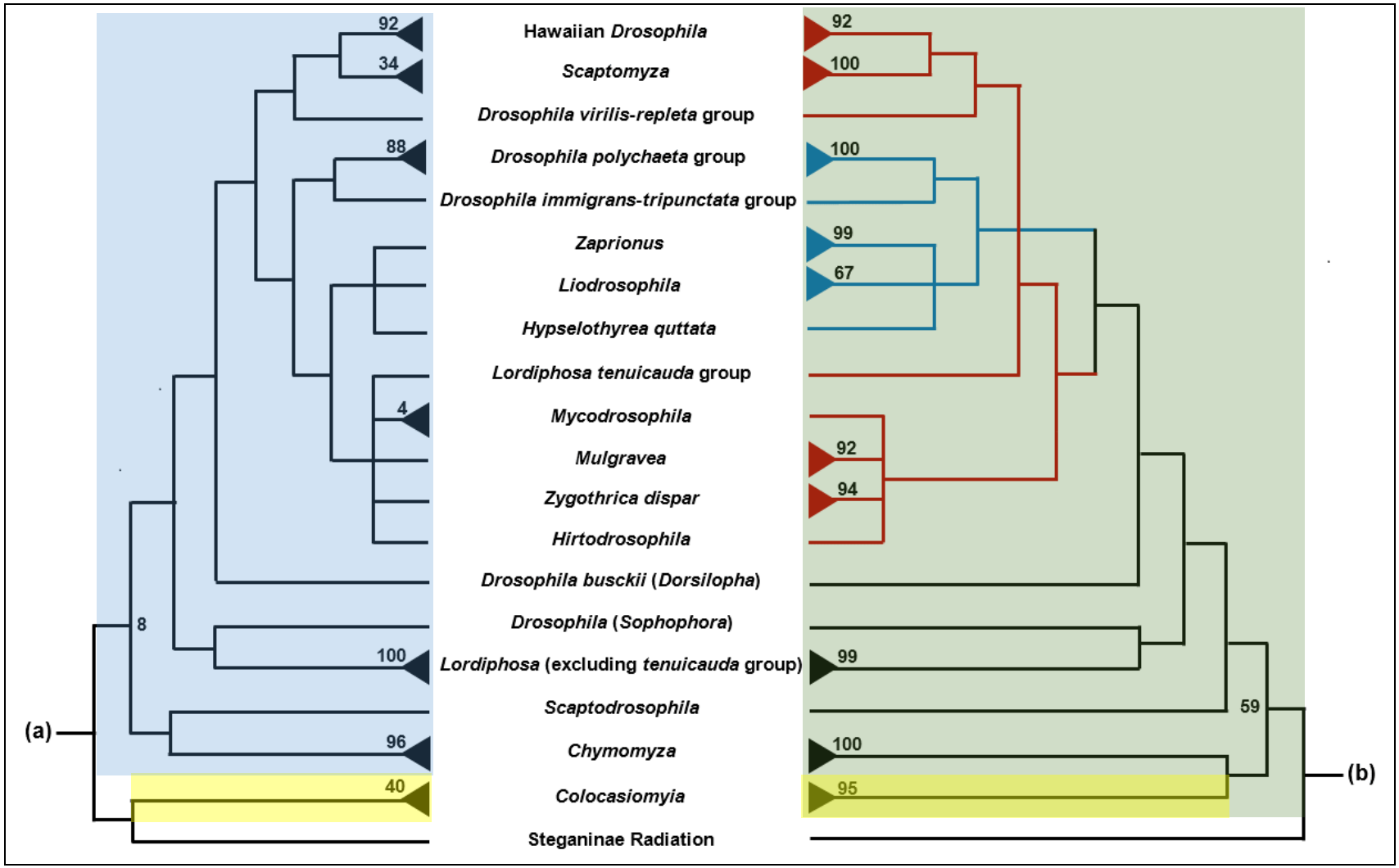

Fig. 52. Drosophilidae (Drosophilinae) phylogeny using the Supermatrix approach after removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. : Drosophilinae, : Drosophilina, : Colocasiomyina.

4.3.4 Gblocks Alignment Masking

The supermatrix created from the concatenation of protein encoding genes and Gblocks-masked rRNA genes contained 1,086 species with 30,130 columns of nucleotide characters containing 87.4% missing characters (online supplementary materials SM19 at https://goo.gl/soTHJk). The RogueNaRok pipeline found 210 rogue species (including *Acletoxenus formosus* and *Leucophenga brevivena*) with 156 taxa that were also detected in the unmasked supermatrix approach (online supplementary material SM15 at https://goo.gl/soTHJk). After removing 210 rogue taxa, the supermatrix contained 8,77 species with 30,130 columns of nucleotide characters containing 86.1% missing characters (online supplementary materials SM20 at https://goo.gl/soTHJk). A summary of the resulting phylogenies by the MP and ML approaches were presented in Fig. 53 and 54 while the full trees can be found in the online supplementary materials SM21 and SM22 at https://goo.gl/soTHJk. Fig. 53 shows the relationships in the Steganinae radiation while Fig. 54 shows a summary of where most individuals of a clade in the Drosophilinae radiation was found.

Overall, there were a few conflicts in relationships generated between the MP and ML approaches (Fig. 53 & 54). Drosophilidae was monophyletic in both trees generated (MP bootstrap = 49%, ML bootstrap = 92%). At the subfamily level, the MP analysis showed Drosophilinae and Steganinae to be paraphyletic. This was because *Colocasiomyia* (Drosophilidae) was nested with a Steganinae, *Phortica* (bootstrap = 2%; Fig. 53) while *Leucophenga saigusai* (Steganinae) and *Leucophenga ornata* (Steganinae) were nested in the *Drosophila virilis-repeleta* species group clade from Drosophilinae (bootstrap = 2%; Fig. 53). The

ML tree showed Steganinae was not monophyletic as the monophyletic Drosophilinae (bootstrap = 49%) was nested within Steganinae (Fig. 53) with a low bootstrap support of 16%.

Only one subtribe, Colocasiomyina (*Colocasiomyia*) was monophyletic (MP bootstrap = 53%, ML bootstrap = 96%; Fig. 53). The Steganinae Gitonini subtribe Gitonina were all clustered together with the exception of *Gitona pauliani* (Fig. 53) *Rhinoleucophenga* and *Acletoxenus,* which were classified under the Steganinae Gitonini subtribe, Acletoxenina, did not group together (Fig. 53). The higher relationships between the genera in Drosophilinae were different for the clade consisting of (*Drosophila polychaeta* species group + *Drosophila immigrans-tripunctata* species group), (*Lordiphosa* tenuicauda species group + *Zaprionus* + *Liodrosophila area* + *Hypselothyrea guttata*), (*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*) and (*Drosophila virilis-repleta* + *Scaptomyza* + Hawaiian *Drosophila*) (Fig. 54). In the MP tree, (*Lordiphosa tenuicauda* species group + *Zaprionus* + *Liodrosophila area*. + *Hypselothyrea guttata*) formed a sister clade to (*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*). These two sister clades were into a sister clade to the clade that contained (*Drosophila polychaeta* species group + *Drosophila immigrans-tripunctata* species group). The clade that contained all the previously mentioned clades formed a sister clade to the clade of [*Drosophila virilis-repleta* species group, (*Scaptomyza* + Hawaiian *Drosophila*)] The ML tree in contrast has (*Drosophila polychaeta* species group + *Drosophila immigrans-tripunctata* species group) as a sister clade to the clade of *(Lordiphosa tenuicauda* species group + *Zaprionus* + *Liodrosophila area* + *Hypselothyrea guttata*). The two sister clades formed a

sister group to the clade consisting of {(*Zygothrica dispar* + *Mulgravea* + *Mycodrosophila* + *Hirtodrosophila*), [*Drosophila virilis-repleta* species group, (*Scaptomyza* + Hawaiian *Drosophila*)]}.

The Steganinae genera were generally monophyletic. The exceptions included the *Leucophenga abbreviata* species group (consisting of *Leucophenga zhenfangae, Leucophenga abbreviata* and *Leucophenga sujuanae*), *Stegana mehadiae* and *Pseudostegana bilobata* in both analyses, in addition to *Leucophenga saigusai*, *Leucophenga ornata* and *Phortica picta* in the MP analysis (Fig. 53). The following clades were monophyletic in both trees (Fig 53 & 54): *Acletoxenus* (MP bootstrap = 86%, ML bootstrap = 100%), *Parastegana* (MP bootstrap = 94%, Rog bootstrap = 100%), *Colocasiomyia* (MP bootstrap = 53%, ML bootstrap = 96%), *Chymomyza* (MP bootstrap = 87%, ML bootstrap = 100%), *Drosophila polychaeta* species group (MP bootstrap = 86%, ML bootstrap = 100%), *Scaptomyza* (MP bootstrap = 46%, ML bootstrap = 99%) and the Hawaiian *Drosophila* species group (MP bootstrap = 92%, ML bootstrap = 95%).

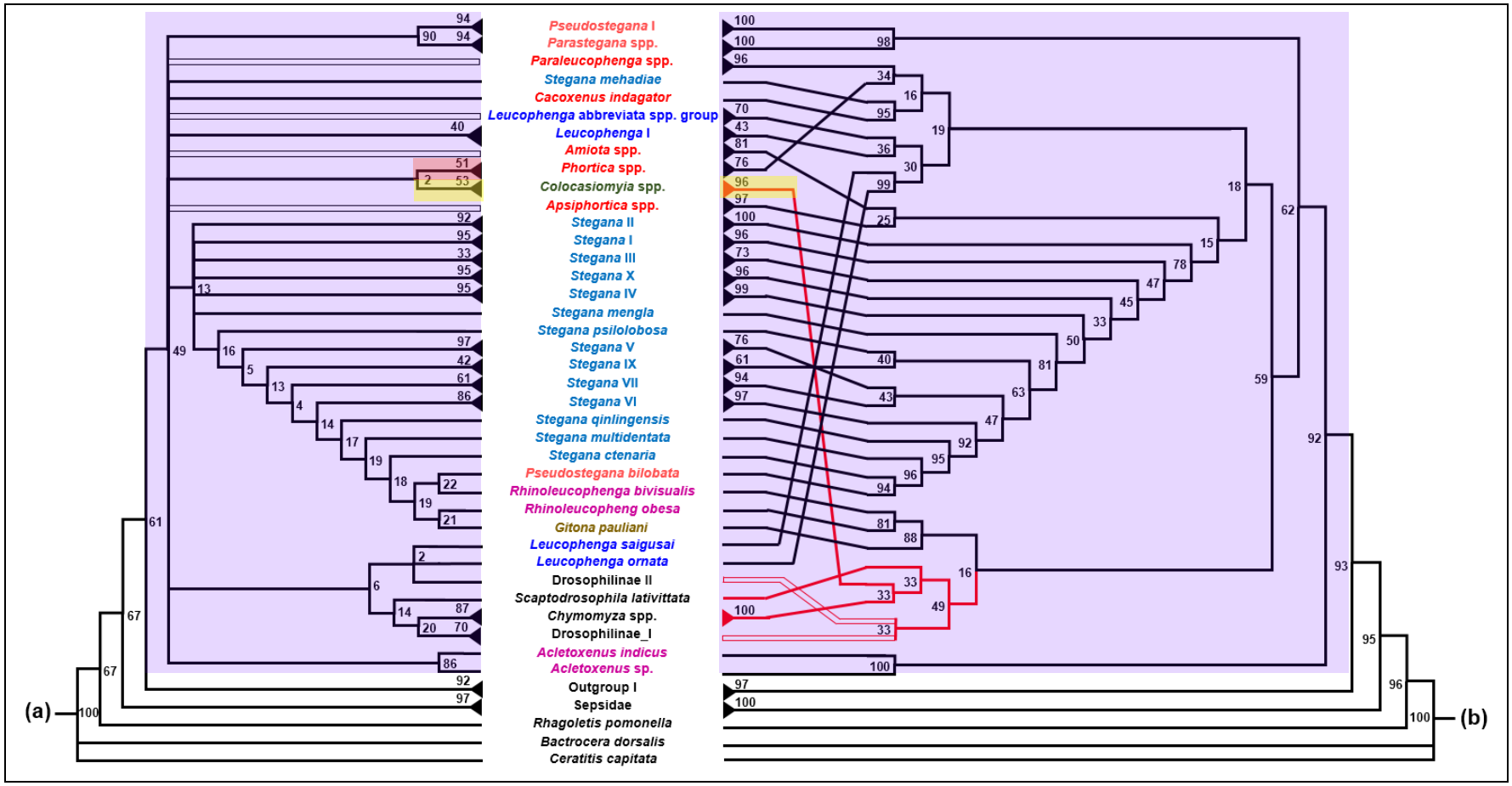

Fig. 53. Drosophilidae (Steganinae) phylogeny using the Supermatrix approach after alignment masking with Gblocks and removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. ▭: Unresolved collapsed phylogeny, ▭: Drosophilidae, ▬: Drosophilinae, ▭: Colocasiomyina, ▭: *Phortica*.

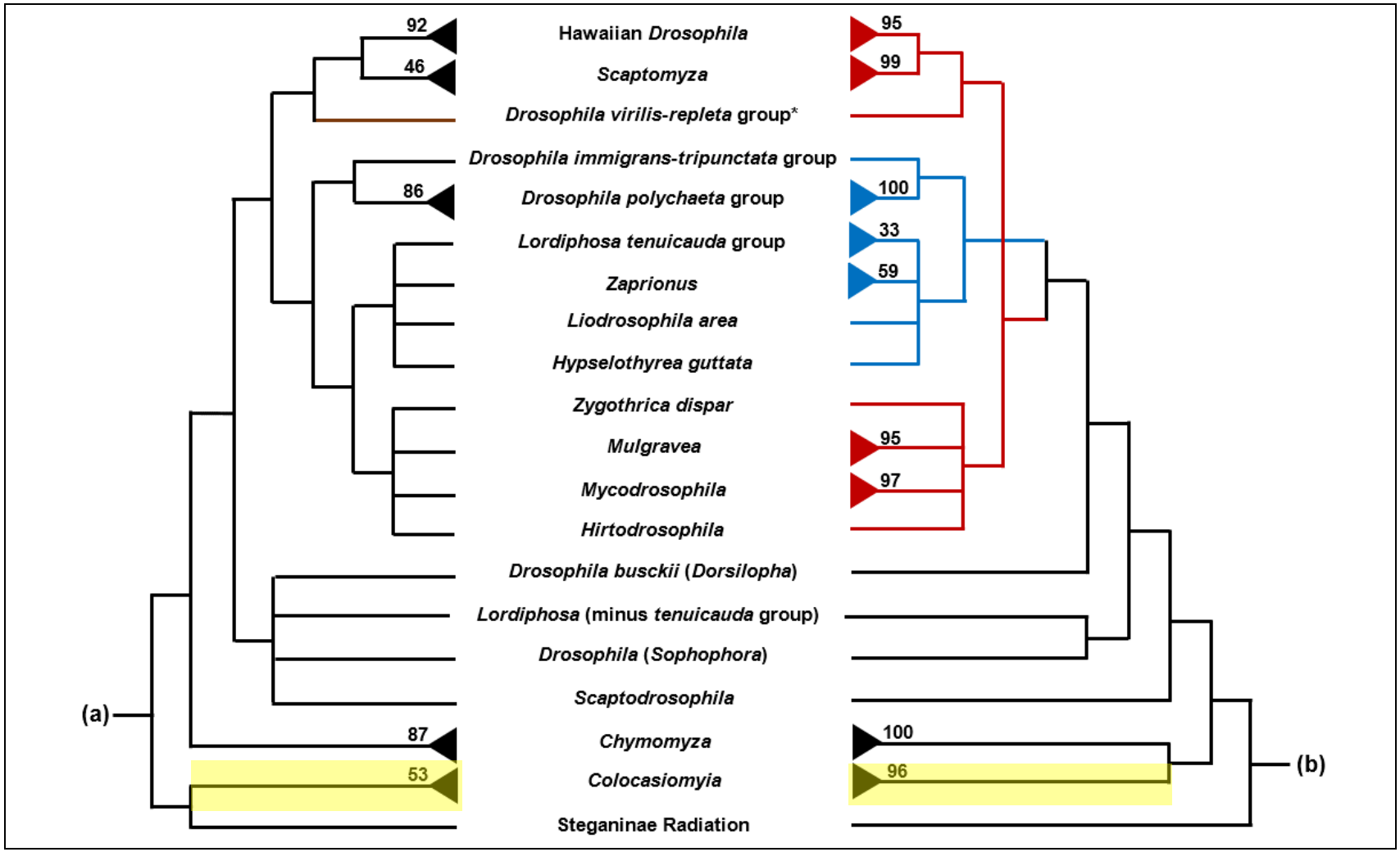

Fig. 54. Drosophilidae (Drosophilinae) phylogeny using the Supermatrix approach after alignment masking with Gblocks and removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. **Leucophenga saigusai* and *Leucophenga ornata* were nested in *Drosophila* (virilis-repleta). : Colocasiomyina.

4.3.5 Zorro Alignment Masking

The supermatrix created from the concatenation of protein encoding genes and Zorro-masked rRNA genes contained 1,086 species with 30,493 columns of nucleotide characters containing 87.5% missing characters (online supplementary materials SM23 at https://goo.gl/soTHJk). The RogueNaRok pipeline found the same 207 species (including *Acletoxenus formosus* and *Leucophenga zhenfangae*) that were identified to be unstable in the unmasked supermatrix (online supplementary materials SM15 at https://goo.gl/soTHJk). After removing the rogue taxa, the supermatrix contained 879 species with 30,493 columns of nucleotide characters containing 86.3% missing characters (online supplementary materials SM24 at https://goo.gl/soTHJk). A summary of the resulting phylogenies by the MP and ML approaches were presented in Fig. 55 and 56 while the full trees can be found in the online supplementary materials SM25 and SM26 at https://goo.gl/soTHJk. Fig. 55 shows the relationships in the Steganinae radiation while Fig. 56 shows a summary of where the most individuals of a clade in the Drosophilinae radiation was found.

Overall, there were a few conflicts in relationships generated between the MP and ML approaches (Fig. 55 & 56). However, Drosophilidae was monophyletic in both trees generated (MP bootstrap = 37%, ML bootstrap = 80%). At the subfamily level, Steganinae was not monophyletic in both tree. In the MP tree, *Colocasiomyia* (Drosophilinae) was nested with *Acletoxenus*, which was classified under Steganinae, although the bootstrap support values were quite low at 3%. In the ML supertree, the monophyletic Drosophilinae (bootstrap = 71%) was nested within Steganinae with a low bootstrap value of 18%.

Only one subtribe, Colocasiomyina (*Colocasiomyia*) was monophyletic (MP bootstrap = 50%, ML bootstrap = 96%; Fig. 55). The Steganinae Gitonini subtribe Gitonina (which includes *Amiota*, *Apsiphortica*, *Cacoxenus*, *Gitona*, *Paraleucophenga* and *Phortica*.) were all clustered together with the exception of *Gitona pauliani* (Fig. 55). *Rhinoleucophenga* and *Acletoxenus* which were classified under the Steganinae Gitonini subtribe Acletoxenina did not group together (Fig. 55). The higher relationships between the Drosophilinae genera were similar to that described in the Supermatrix analyses in Section 4.2.3.

The Steganinae genera were generally nested together in both trees with the exception of *Pseudostegana bilobata*, *Stegana mehadiae*, *Leucophenga abbreviata* species group (consisting of *Leucophenga brevivena*, *Leucophenga abbreviata* and *Leucophenga sujuanae*) in both trees and *Phortica picta* in the MP tree (Fig. 55). The following clades below the genera level were monophyletic in both trees (Fig 55 & 56): *Acletoxenus* (MP bootstrap = 84%, ML bootstrap = 100%), *Apsiphortica* (MP bootstrap = 52%, Rog bootstrap = 98%), *Parastegana* (MP bootstrap = 97%, Rog bootstrap = 100%), *Colocasiomyia* (MP bootstrap = 50%, ML bootstrap = 96%), *Drosophila polychaeta* species group (MP bootstrap = 90%, ML bootstrap = 100%), *Scaptomyza* (MP bootstrap = 46%, ML bootstrap = 99%) and the Hawaiian *Drosophila* species group (MP bootstrap = 92%, ML bootstrap = 93%). *Acletoxenus* (*Acletoxenus indicus* and *Acletoxenus* cf. *indicus*) was monophyletic in both trees (MP bootstrap = 84%, ML bootstrap = 100%).

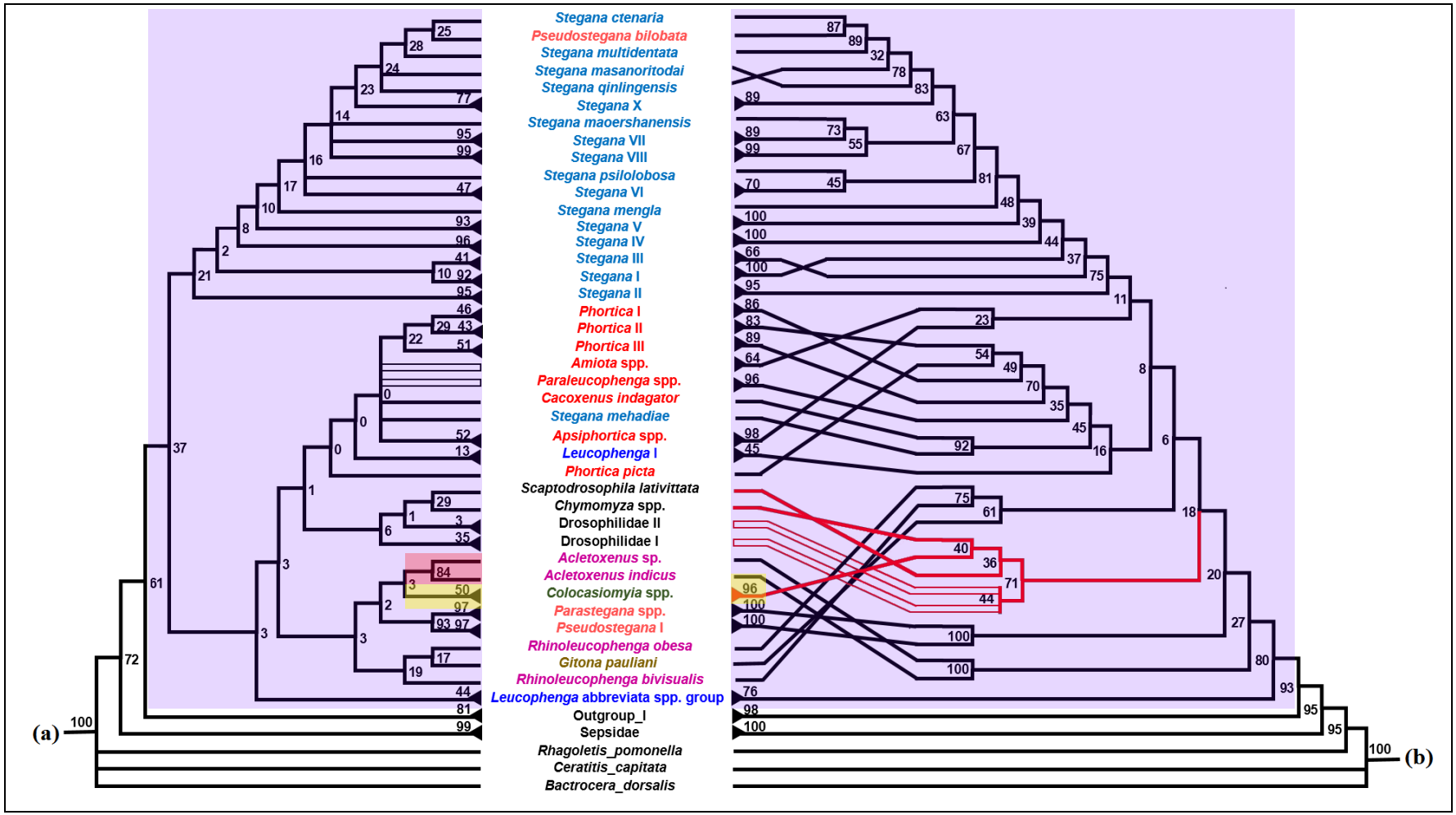

Fig. 55. Drosophilidae (Steganinae) phylogeny using the Supermatrix approach after alignment masking with Zorro and removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. ▭: Unresolved collapsed phylogeny, ▭ : Drosophilidae, ▭ : Drosophilinae, ▭ : Colocasiomyina, ▭ : *Acletoxenus*.

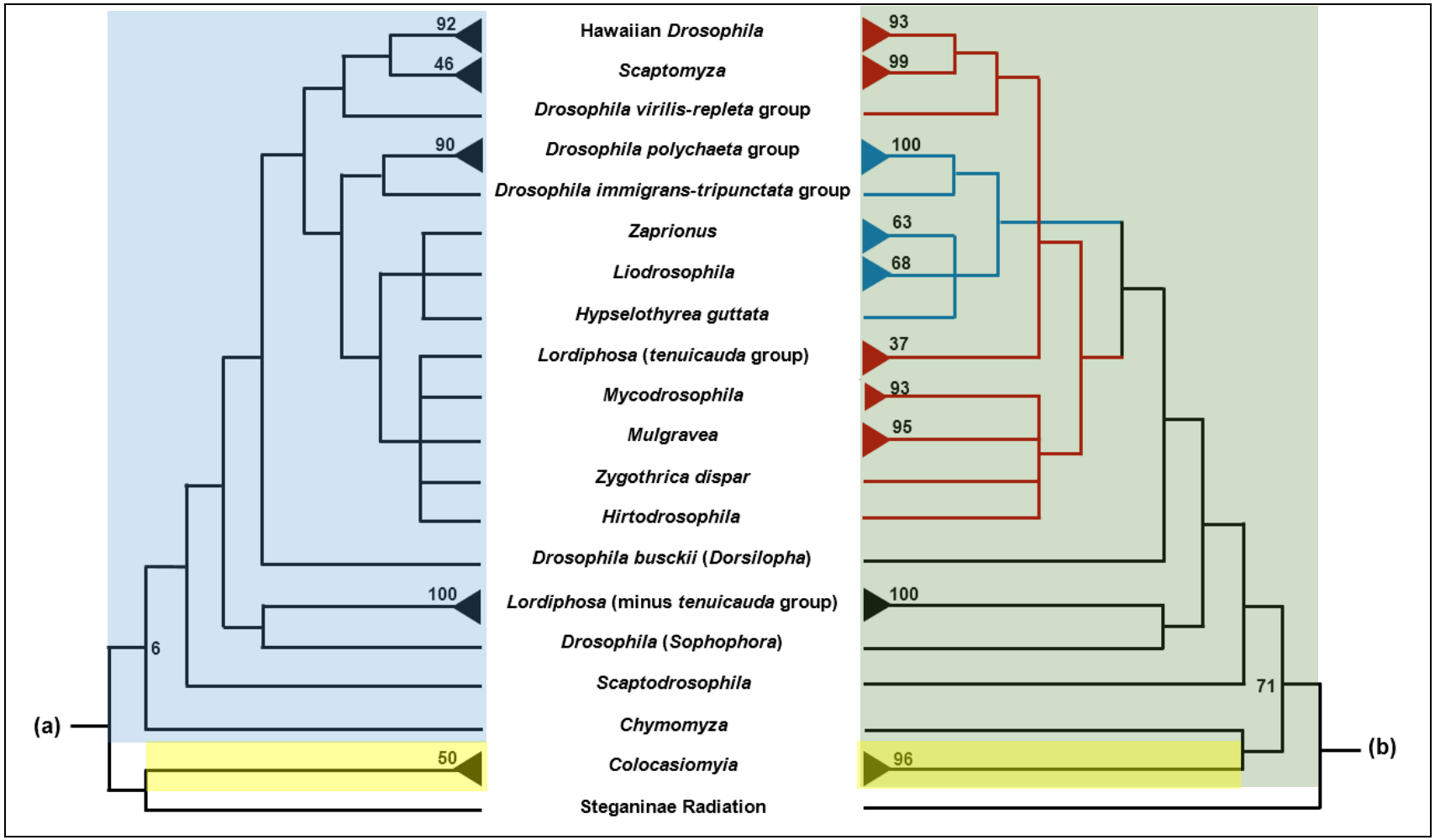

Fig. 56. Drosophilidae (Drosophilinae) phylogeny using the Supermatrix approach after alignment masking with Zorro and removal of rogue taxa by RogueNaRok using (a) MP and (b) ML. : Drosophilinae, : Drosophilina, : Colocasiomyina.

4.3.7 Post-Phylogenetic Analysis Rogue Taxa Check

The full rogue taxa list generated by IterPCR and RogueNaRok are in the online supplementary materials SM23 at https://goo.gl/soTHJk. The summary of whether a species (that was not nested with individuals of its classification) was a rogue taxa is in Table 4. Only *Gitona pualiani* and *Pseudostegana bilobata* were not identified as rogue taxa while the other species were identified as rogue taxa in one of the analyses (Table 4). The mean interspecific pairwise distance in *Pseudostegana* was 12.76% in COI, 15.29% in ND2, 0.70% in 28Sd1, 3.25% in 28Sd2 and 2.84% in 28Sd3.

Table 4. Summary of whether a species that was nested with the rest of its clade was a rogue taxon.

Species not nested with the rest of its clade	Unmasked Supermatrix MP IterPCR	Unmasked Supermatrix ML RogueNaRok	Gblocks Supermatrix MP IterPCR	Gblocks Supermatrix ML RogueNaRok	Zorro Supermatrix MP IterPCR	Zorro Supermatrix ML RogueNaRok
Gitona pualiani	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected
Leucophenga brevivena	Not detected	Not detected	Rogue taxon*	Rogue taxon*	Not detected	Not detected
Leucophenga abbreviata	Not detected	Not detected	**Rogue taxon**	Not detected	Not detected	Not detected
Leucophenga sujuanae	Not detected	Not detected	**Rogue taxon**	Not detected	Not detected	Not detected
Leucophenga zhenfangae	Rogue taxon*	Rogue taxon*	**Rogue taxon**	**Rogue taxon**	Rogue taxon*	Rogue taxon*
Pseudostegana bilobata	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected
Stegana mehadiae	**Rogue taxon**	Not detected	**Rogue taxon**	Not detected	**Rogue taxon**	Not detected
Phortica picta	Not detected	**Rogue taxon**	Not detected	Not detected	Not detected	**Rogue taxon**
Lordiphosa acutissima	**Rogue taxon**	**Rogue taxon**	Not detected	**Rogue taxon**	**Rogue taxon**	**Rogue taxon**
Lordiphosa pseudotenuicauda	Not detected	Not detected	**Rogue taxon**	Not detected	Not detected	Not detected
Lordiphosa tenuicauda	Not detected	Not detected	**Rogue taxon**	Not detected	Not detected	Not detected

* These rogue taxon were previously detected in the first round of rogue taxa check.

4.3.8 The Position of *Acletoxenus* In The Supermatrix Phylogenetic Trees

On the supertree, *Acletoxenus* is a sister clade to the other Drosophilidae on the Concatabomination ML tree (Fig. 47) but nested with *Teleopsis whitei* in the RogueNaRok ML tree (Fig. 48). In the unmasked and Zorro-masked supermatrix ML trees, *Acletoxenus* formed a sister clade to the other Drosophilidae excluding the *Leucophenga abbreviata* species group while in the MP tree, Acletoxenus was a sister clade to *Colocasiomyia* and nested with the other Gitonini (Fig. 51, 55 & 57). In the Gblocks-masked supermatrix analysis, *Acletoxenus* formed a sister clade to the other Drosophilidae in the ML tree but was unresolved in the MP tree (Fig. 53 & 57).

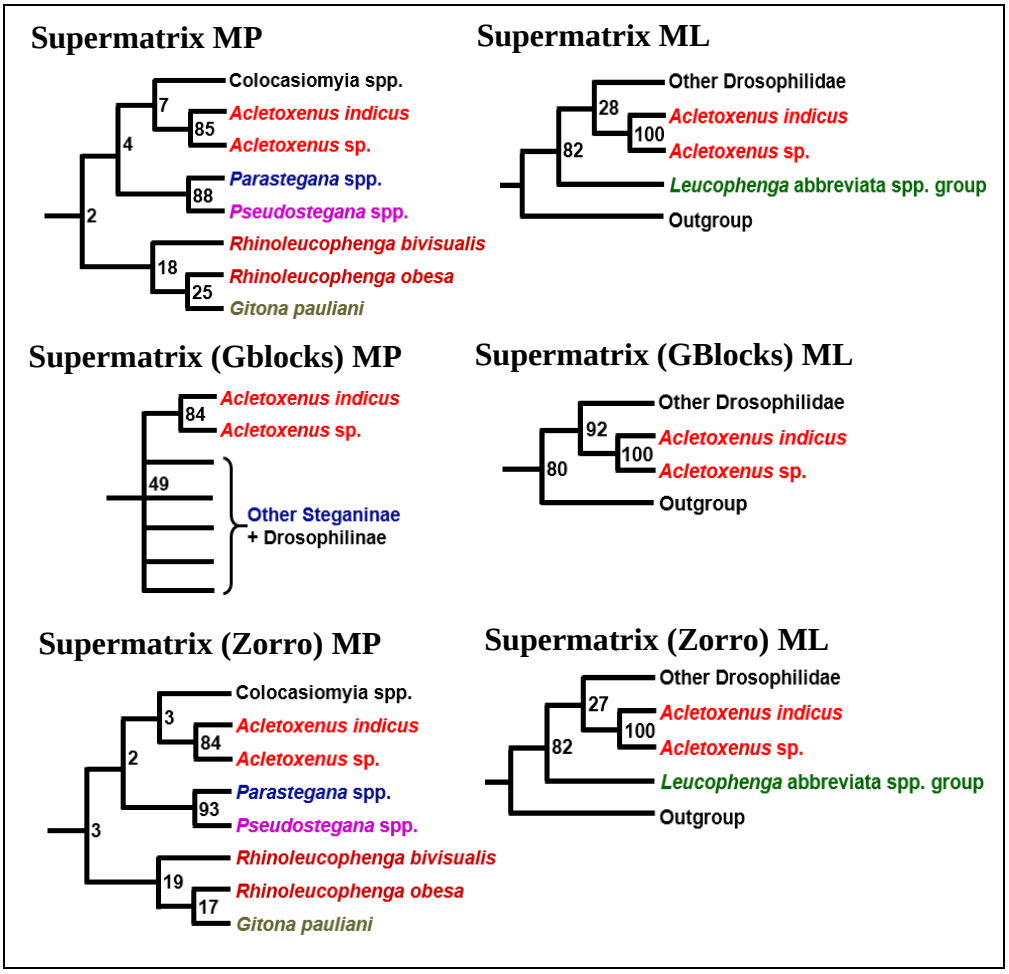

Fig. 57. The position of *Acletoxenus* in the supermatrix phylogenetic trees.

4.4 Discussion

4.4.1 Overall Comparison

Overall, the Drosophilidae phylogeny cannot be resolved with the available data. This was because the topologies between the clades above the genera level were not congruent between the various analyses (Fig. 49 – 56). This could have arisen due to different alignment treatments (MRP in supertree as compared to concatenation in supermatrix) as well as the different optimality criterion (MP and ML) used to build the phylogeny.

The contrasting relationships between the previous and current analyses could be a result of the current analyses having a bigger set of data and which used deeper analysis algorithms than previous studies. Although many of the genera were monophyletic and well supported with high bootstrap values in this study, the higher relationships above the genera level had very low support values and had topologies that were not congruent in the different analyses. This could be a result of the high number of missing genes such that were insufficient defining characters to differentiate the genera.

In all the analyses, most members of a genus would cluster together except for a few species. These few species were represented by a small number of genes and were most likely rogue taxa although they were not identified by the rogue taxa programs. This indicated that the current programs used to identify rogue taxa (RogueNaRok, Concatabomination and IterPCR) were unable to fulfil their intended functions. Furthermore, rogue taxa have been known to impact topological resolution negatively, which might have resulted in the lack of

congruence between the analyses (Sanderson *et al.*, 1998; Wilkinson, 1995, 2003).

As the incongruent and lowly supported topologies plus rogue taxa were a consequence of a big number of missing genes in the matrices used to generate the phylogeny, even though GenBank currently has a huge amount of data, it is still insufficient for obtaining a well-resolved and supported phylogenetic hypothesis for Drosophilidae. Hence, more genes would have to be obtained to enable one to obtain a well-resolved and supported phylogenetic hypothesis for Drosophilidae.

4.4.2 Supertree & Supermatrix Approach Comparison

4.4.2.1 Alignment Treatment

As the MP method in the supermatrix approach was able to resolve more relationships within Drosophilidae, the use of the supermatrix over the supertree method for MP was better. This occurred as the characters used to define each clade in the alignments were lost when they were converted into the binary MRP in the supertree method (de Queiroz & Gatesy, 2007). In contrast, the supertree ML analysis may produce an accurate tree since the topologies and the monophyletic clades in both ML analyses were similar and well supported.

4.4.2.2 Concatabomination and RogueNaRok Comparisons

In the ML analyses, both RogueNaRok and Concatabomination produced well resolved trees as a consequence of the ML approach. However, the monophyletic clades in the RogueNaRok best tree had higher support values than the Concatabomination best tree (Fig. 49 & 50). This suggested that RogueNaRok was better at identifying rogue taxa than Concatabomination. On

the other hand, Concatabomination did a better job at identifying rogue taxa as it resulted in slightly better resolved phylogenies as compared to RogueNaRok in the MP analyses (Fig. 49 & 50).

The difference in rogue taxa detected between Concatabomination and RogueNaRok was a result of the different algorithms used as Concatabomination is an *a priori* method for looking for unstable taxa while RogueNaRok is an *a posteriori* method (Fig. 49 & 50). The lack of resolution in the MP RogueNaRok analysis could be because the assumption that the rogue taxa detected by the RogueNaRok algorithm from the ML analysis would be similar for the MP analysis was false. Then again, it could be that the poor resolution could be a result of the supertree MP approach as regardless of whether Concatabomination or RogueNaRok was used to remove rogue taxa, this approach was unable to resolve any relationships well (Fig. 49 & 50).

4.4.2.3 MP & ML Comparisons

Between the two different optimality criteria, ML was able to produce a better resolved tree than MP in the supertree method. As the topologies in the ML analyses with Concatabomination and RogueNaRok were similar in addition to monophyletic clades that were well supported, the supertree ML analysis may produce an accurate analysis. However, one should be cautioned of the results as the same MRP used in the MP analyses was unable to resolve much relationships.

4.4.3.4 Similar Topologies

There were some concordance in relationships between the supertree and supermatrix approaches which gives us more confidence that these relationships

are likely correct. This included the basal radiation of *Scaptodrosophila* and *Chymomyza* in Drosophilinae and the sister group of the monophyletic Hawaiian Drosophilidae clade (Hawaiian *Drosophila* species group and *Scaptomyza*) and the *Drosophila virilis-repleta* species group which agreed with previous phylogenetic analyses (DeSalle, 1992; DeSalle & Grimaldi, 1991; Remsen & O'Grady, 2002; van der Linde *et al.*, 2010; Yassin, 2013). *Scaptodrosophila* and *Chymomyza* were not monophyletic in most analyses which is also in concordance with recent analyses (van der Linde *et al.*, 2010; Yassin, 2013) while older publications still suggested that the two genera were monophyletic and that *Scaptodrosophila* diverged from the other drosophilids before *Chymomyza* (DeSalle, 1992; Hu & Toda, 2001; Kwiatowski *et al.*, 1997; Kwiatowski *et al.*, 1994; Tarrío *et al.*, 2001).

4.4.3 Supermatrix Approach With Alignment Masking

4.4.3.1 Zorro & Gblocks Comparison

The RogueNaRok program found the same number of rouge taxa (207) from the unmasked and Zorro-masked supermatrix ML bootstraps and 209 rogue taxa based on the Gblocks-masked supermatrix ML bootstraps. This suggested that Gblocks may not have improved alignment quality as the number of rogue taxa increased. Similarly, Zorro may not have helped to improve alignment quality since the number of rogue taxa detected was identical. However, this might have been a result of the parameters used for Gblocks and Zorro to detect poorly aligned regions.

Gblocks produced a MP tree that was unresolved for Drosophilidae; i.e., the algorithm in Gblocks was masking defining characters that differentiate these

clades. The results also indicated that important evolution information that may help to resolve the backbone of the Drosophilidae tree was found in the rRNA genes and that these divergent regions were determined to be poorly aligned by Gblocks. The parameters used for Zorro resulted in a phylogenetic tree that had better relationships that were similar to the unmasked supermatrix tree in Drosophilinae but different in Steganinae. Thus, it is conceivable that the relaxed criteria for Zorro may have removed so little low quality regions that it had little effect. In addition, the similar relationships in Drosophilinae but not in Steganinae may be a result of insufficient data for Steganinae, which was also reflected by the low bootstrap values.

As there was not much difference in support values for the monophyletic clades between the unmasked and masked supermatrix trees, the alignment quality of the gene fasta files was generally good as there were no improvements after removal of the poorly aligned regions with either programs. Instead, the low support values were most likely a result of a huge amount of missing data in the supermatrices. Alternatively, the relaxed parameters used in the analyses may not have been able to identify poorly aligned regions as the default parameters were not used. However, the default parameters resulted in largely unresolved phylogenetic trees during a trial test. Thus, the parameters used for alignment masking have a big impact in the phylogenetic analyses and the parameters that allow for identifying poorly aligned regions requires more research.

4.4.3.2 Analysis at the Subfamily level

Drosophilidae was monophyletic in all the trees generated by the supermatrix method although the topologies of the tree did not correspond to the two

subfamilies, Steganinae and Drosophilinae. The supermatrix ML analyses found Drosophilinae to be monophyletic with support values from 49% to 71%. However, Drosophilinae was nested within Steganinae instead of being a sister group as suggested by Yassin's (2013) analysis. The MP analyses also found *Colocasiomyia* (Drosophilinae) to be nested with other Steganinae species, including *Acletoxenus* in the unmasked and Zorro-masked supermatrix analyses (Fig. 51 & 55) and *Phortica* in the Gblocks-masked supermatrix analysis (Fig. 53). The placement of *Colocasiomyia* with *Acletoxenus* could have arisen as a result of long-branch attraction, which is a known problem of MP, but also occurs in ML. As *Acletoxenus* was found to be a basal radiation on the ML trees while *Colocasiomyia* belonged to a different subtribe from the other drosophilids in this study, the MP algorithm grouped them together because they had a bigger number of divergent sequences from the other drosophilids, rather than because they are related by descent (Bergsten, 2005). In addition, the bootstrap values that placed *Colocasiomyia* as a sister group to *Acletoxenus* (3% in supermatrix MP tree and 7% in Zorro-masked supermatrix MP tree) or *Phortica* (2% in Gblocks-masked supermatrix MP tree) were very low. This suggested that its current position on the MP trees were most unlikely. Indeed, although *Colocasiomyia* was found to polyphyletic, majority of its members were nested in Drosophilinae in comparison to the current finding where it is nested with *Acletoxenus* (Sultana *et al.*, 2006).

Although previous analyses showed that the subfamilies Steganinae and Drosophilinae were monophyletic clades, the results from this study suggests otherwise. This difference in results is significant because previous analyses only used a small number of Steganinae in their analyses; thus effectively only

treating them as outgroups. The paper with the largest number of Steganinae is Yassin (2013) who used only 23 species. Thus, the current classification and proposed basal division into the subfamilies requires more scrutiny.

4.4.3.3 Analysis at the Tribe & Subtribe Level

None of the phylogenetic trees displayed topologies that were in line to the tribe- and subtribe-level classification except for Colocasiomyina (Fig. 51, 53 & 55). In the ML trees, Colocasiomyina was monophyletic and nested in Drosophilina (Fig. 51, 53 & 55). This was in contrast to the study by Sultana *et al.* (2006) which found *Colocasiomyia* to be paraphyletic. The reason for this is because the current study did not include species from the *Colocasiomyia arenga* species group that were found to be separated from the other *Colocasiomyia* species. Thus, the *Colocasiomyia arenga* species group would have to be included to get a better conclusion on the status of this subtribe.

4.4.3.4 Analysis of the Relationships between Genera Level

There was only one relationship between the Steganinae genera that was in concordance to previous phylogenetic analyses; the group of (*Apsiphortica* + *Phortica* + *Cacoxenus indigator*) being nested together in the unmasked and Zorro-masked supermatrix MP analyses were similar to the findings of Yassin (2013) (Fig. 51 & 55). The lack of concordance in the other relationships between the Steganinae genera may be a result of lack of data (Fig. 51, 53 & 55). More data would be required before one can make a better judgement of the actual positions of the Steganinae genera.

On the other hand, the relationships between the Drosophilinae genera showed more concordance with previous phylogenetic analyses (Fig. 52, 54 & 56). The

clade of (*Scaptomyza* + Hawaiian *Drosophila* species group) was found in all trees in this study, similar to those of Yassin (2013) and van der Linde *et al.* (2010). The *Drosophila virilis-repleta* species group was found as the sister group to the clade of (*Scaptomyza* + Hawaiian *Drosophila* species group), similar in van der Linde *et al.* (2010). The group consisting of (*Zaprionus* + *Liodrosophila* + *Hypselothyrea guttata*) was similar to that of Yassin (2013) and van der Linde *et al.* (2010). Likewise, the clade of (*Zygothirca dispar* + Mulgravea + Mycodrosophila + Hirtodrosophila) was also present in the phylogenetic analysis by Yassin (2013) and van der Linde *et al.* (2010). Similar to Yassin (2013) and van der Linde *et al.* (2010), *Drosophila busckii* was a sister group to the clade consisting of (*Scaptomyza* + Hawaiian *Drosophila* species group+ *Drosophila virilis-repleta* species group + *Drosophila polychaeta* species group + *Zaprionus* + *Liodrosophila* + *Hypselothyrea guttata*) in all but the Gblocks-masked supermatrix MP analysis.

4.4.3.5 Analysis at the Steganinae Genus & Species Group Level

Gitona pauliani was not nested with the members of Gitonini in all the supermatrix analyses. Since it was not detected as a rogue taxa, its position of being nested in the *Rhinoleucophenga* clade raised the possibility that *Gitona pauliani* might be from the *Rhinoleucophenga* genus (Fig. 51, 53 & 55; Table 4). Over the years, many authors have pointed out many *Rhinoleucophenga* species (including *Rhinoleucophenga bivisualis*) that were previously classified as *Gitona* to be incorrectly placed (Ashburner, 1981; Bächli *et al.*, 2004; McAlpine, 1968; Otranto *et al.*, 2008; Wheeler & Takada, 1971). This was because the traditional circumscription of *Rhinoleucophenga* was the presence of plumose aristae (Schmitz, 2010). However, there were other species that had

bare or micropubescent aristae which had characteristics of *Rhinoleucophenga* as well. Thus, one should re-examine the specimens of *Gitona pauliani* that was described in 1951 to better determine if its genus should be revised (Séguy, 1951).

The *Leucophenga abbreviata* species group was recently proposed in 2013, and consisted of *Leucophenga abbreviata*, *Leucophenga brevivena*, *Leucophenga sujuanae* and *Leucophenga zhenfangae* (Su *et al.*, 2013). In the unmasked and Zorro-masked supermatrix phylogenetic analyses, *Leucophenga zhenfangae* was found to be a rogue taxa by RogueNaRok before the phylogenetic analysis. After its removal, the remaining three species formed an outgroup to the other species of Drosophilidae (Fig. 51 & 55). In the Gblocks-masked supermatrix analysis, *Leucophenga brevivena* found to be a rogue taxa by RogueNaRok before the phylogenetic analysis. After its removal, the three remaining species in the group nested with the other *Leucophenga* species. (Fig. 53). Thus, it could be that *Leucophenga brevivena* was a rogue taxa that was undetected by RogueNaRok before the phylogenetic analysis in the unmasked and Zorro-masked supermatrix. This was supported by the low bootstrap values in the MP analyses of 1% in the unmasked supermatrix and 3% in the Zorro-masked supermatrix (Fig. 51 & 55). Furthermore, all the species from the *Leucophenga abbreviata* species group had only one gene (COI) used in the analysis. As such, the data may not have had enough common characters to cluster the members of this group with the other *Leucophenga* species. This could be due to the high interspecific distances of COI within *Leucophenga* that were found to be between 8.36% and 11.39% (Su *et al.*, 2013). On the other hand, *Leucophenga brevivena* was still not identified as a rouge taxa in the post-phylogenetic

analysis rogue taxa check (Table 4) and its position as a sister group to the other drosophilids had high support in the ML analyses at 82% in the unmasked supermatrix analysis and 80% in the Zorro-masked supermatrix analysis (Fig. 51 & 55). Thus, more genes should be obtained for the *Leucophenga abbreviata* species in order to identify their position in the phylogeny of Drosophilidae.

As *Pseudostegana bilobata* was not detected as a rogue taxa in the post-phylogenetic analysis, this suggests that its current position is not an error (Table 4). This result of *Pseudostegana bilobata* being found within the *Stegana* clade in all except the Gblocks-masked MP analysis was not surprising (Fig. 51, 53 & 55). This was because *Pseudostegana* was formerly a subgenus of *Stegana* before it was elevated to genus level by Sidorenko (2002). Thus, these two closely related genera shared many common features, including the absence of prescutellar setae, ocellar setae being situated outside of the ocellar triangle, presence of one pair of dorsocentral setae and the wing vein M being strongly convergent to R_{4+5} (Chen *et al.*, 2005; Sidorenko, 2002). In the Gblocks-masked MP analysis as well, *Pseudostegana bilobata* was not nested with the members of its own genus but nested with *Rhinoleucophenga bivisualis* instead (Fig. 53). This position of *Pseudostegana bilobata* which is not nested within the remaining members of its genus could have been a result of the huge interspecific distance in COI (pairwise distance = 12.76%) and ND2 (pairwise distance = 15.29%) genes, such that there was insufficient common characters to group them together. In addition, there was also a contrast in support values for the current position of *Pseudostegana bilobata*, being low in the MP approach (bootstrap = 22% to 25%) but high in the ML approach (bootstrap = 87% to 94%). On the other hand, the 28S domain genes had much lower

interspecific distance of 0.70% in 28Sd1, 3.25% in 28Sd2 and 2.84% in 27Sd3. Thus, additional genes would be required to better represent *Pseudostegana* for a more accurate phylogenetic placement.

Instead of being nested with the other *Stegana* species, *Stegana mehadiae* was nested with *Cacoxenus indagator* with a high support of 90% to 95% in the ML analyses although it was unresolved in the MP analyses (Fig. 51, 53 & 55). This same relationship was also present in the COI tree by Otranto *et al*. (2008). As *Stegana mehadiae* was only characterized by COI, it may not have contained enough common characters to group it with the other *Stegana* species. Indeed, its relationship was unresolved, had no support in the MP analyses and was detected as a rogue taxa in the post-phylogenetic analysis rogue taxa check (Fig. 51, 53 & 55; Table 4).

Phortica picta was not nested with the other *Phortica* species in the unmasked and Zorro-masked supermatrix phylogenetic MP analyses (Fig. 51 & 55). As *Phortica picta* was picked up as a rouge taxa in the Gblocks-masked supermatrix analysis, it could be that *Phortica picta* was a rogue taxa that was undetected by RogueNaRok before the phylogenetic analysis in the unmasked and Zorro-masked supermatrix. Indeed, the post-phylogenetic analysis found *Phortica picta* to be a rogue taxa (Table 4). This was supported by the lack of support for its current position in the unmasked (bootstrap = 2%) and Zorro-masked (bootstrap = 0%) supermatrix analyses that indicated its placement was most likely incorrect. However, in the unmasked and Zorro-masked supermatrix ML analyses, *Phortica picta* was nested with the other *Phortica* species such that the clade was monophyletic with a reasonable support of 70% and 77% (Fig. 51 & 55). Thus, although *Phortica picta* was most likely nested with the other

Phortica species, it would be best to re-run the analysis using more gene coverage for all the *Phortica* species as a majority of them were only represented by COI in the analyses.

4.4.3.6 Analysis at the Drosophilinae Genus & Species Group Level

Scaptomyza and Hawaiian *Drosophila* species group were monophyletic clades in all the analyses (Fig. 52, 54 & 56). Similarly, most of the individuals in the Drosophilinae genera, *Zaprionus*, *Liodrosophila*, *Hypselothyrea, Mulgravea, Mycodrosophila* and each group of *Drosophila* were nested together with their respective groups. *Hirtodrosophila* formed a clade with *Mulgravea, Mycodrosophila* and *Zygothrica* although it was paraphyletic and grouped with individuals of other genera from the clade mentioned (Fig. 52, 54 & 56). This could be because of the small number of genes that overlapped for each species, such that they shared little common characters so they did not nest together. The other genera that was visibly paraphyletic was *Lordiphosa* (Fig. 52, 54 & 56). Most of the species under the genera *Lordiphosa* were nested together and formed a sister group to *Drosophila* (*Sophophora*) except for the species in the *Lordiphosa tenuicauda* species group (Fig. 52 & 56). This group, comprising of *Lordiphosa acutissima, Lordiphosa pseudotenuicauda* and *Lordiphosa tenuicauda,* was nested with (*Mulgravea* + *Mycodrosophila* + *Hirtodrosophila* + *Zygothrica dispar*) in the unmasked and Zorro-masked supermatrix MP analyses (Fig. 52 & 56). This result was similar to that by Katoh *et al*. (2000), where most members of the *Lordiphosa* genera formed a sister group to *Drosophila (Sophophora)* while the *Lordiphosa tenuicauda* species group was nested with *Hirtodrosophila*. On the other hand, the *Lordiphosa tenuicauda* species group was a sister group to (*Drosophila virilis-repleta* species group +

Scaptomyza + Hawaiian *Drosophila*) in the ML analyses (Fig. 52 & 56) and nested with (*Zaprionus* + *Liodrosophila area* + *Hypselothyrea guttata*) in the Gblocks-masked analyses (Fig. 56). This lack of concordance could be a result of the *Lordiphosa tenuicauda* species group having only one gene, ADH, being used in the analyses. Indeed, all members of this group were detected at rogue taxa in at least one of the post-phylogenetic rogue taxa analysis (Table 4). Thus, more genes of these species would have to be added to better determine its position in the Drosophilidae phylogenetic tree.

4.4.4 The Position of *Acletoxenus* in the Phylogenetic Trees

Acletoxenus was found in three different positions in the different phylogenetic analyses: (1) It was a sister group to the other Drosophilidae in the RogueNaRok supermatrix ML analyses; (2) It was nested with *Teleopsis whitei* in the Concatabomination supertree ML analysis; (3) It was nested within Gitonini in the RogueNaRok supertree ML analysis and supermatrix MP analyses. Out of these three positions, the position of being nested within Gitonini was in line with the current classification of *Acletoxenus* although the support values were very low (2% and 3% in unmasked and Zorro-masked supermatrix MP analyses). On the other hand, the support values for the sister group position to the other Drosophilidae was much higher, especially in the Gblocks-masked supermatrix that was at 92%. This position may have arisen because *Acletoxenus* had a much bigger number of gene loci (22) as compared to the other Steganinae genera which had less than 10 gene loci represented in the analysis. Thus, *Acletoxenus* may have been more different from the other Steganinae as there was very little overlap. For the same reasons, *Acletoxenus* may have been placed with *Teleopsis whitei* (outgroup) in the

Concatabomination supertree ML analysis. However, although the unmasked (28%) and Zorro-masked (27%) supermatrix ML analyses that supported the sister group position to the other Drosophilidae had bootstrap support values that were higher than those in the MP analyses, the support values were still very low. Thus, the lack of concordance between the results indicated that the current data was unable to accurately position *Acletoxenus* on the Drosophilidae tree. More gene coverage for the other Steganinae species will be needed to get a better phylogenetic analysis to pinpoint its position.

References

Abascal, F., R. Zardoya & M. J. Telford, 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research* **38**: W7-W13.

Aberer, A. J., D. Krompaß & A. Stamatakis, 2011. RogueNaRok: an efficient and exact algorithm for rogue taxon identification. *Exelixis-RRDR-2011–10*, Heidelberg Institute for Theoretical Studies.

Acletoxenus quadristriatus Duda, 1936. The Atlas of Living Australia website. Updated in 2015. Retrieved on 6 January 2015 from http://bie.ala.org.au/species/urn:lsid:biodiversity.org.au:afd.taxon:f91182e5-676c-4dd0-bba4-8c459980389a.

Agresti, A., & Franklin, C., 2013. *Statistics: The Art and Science of Learning from Data*. 3rd edition. Pearson Higher Ed. 848 pp.

Altincicek, B. & A. Vilcinskas, 2007. Analysis of the immune-inducible transcriptome from microbial stress resistant, rat-tailed maggots of the drone fly *Eristalis tenax*. *BMC Genomics*, **8**: 326.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers & D. J. Lipman, 1990. Basic local alignment search tool. *Journal of molecular biology*, **215**(3): 403-410.

Ashburner, M., 1981. Entomophagous and other bizarre Drosophilidae. In M. Ashburner, Carson, H. L., and Thompson, J. N. (ed.), *The Genetics and Biology of Drosophila.* Academic Press: New York. Vol. 3a, Pp. 375–429.

Ashburner, M. G. & K. Hawley, 2005. *Drosophila: A Laboratory Handbook.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Aubertin, D., 1937. A new species of *Drosophila*, *D. Gibbinsi* sp. N., from Uganda (Diptera). *Proceedings of the Royal Entomological Society of London. Series B, Taxonomy*, **6**(9): 169-169.

Bächli, G., 2015. *TaxoDros: Classification*. TaxoDros, the database on taxonomy of Drosophilidae. Updated on 25 March 2015. Retrieved on 12 April 2015 from http://www.taxodros.uzh.ch.

Bächli, G., C. R. Vilela, S. A. Escher & A. Saura, 2004. *The Drosophilidae (Diptera) of Fennoscandia and Denmark.* Brill Academic Publishers. 362 pp.

Basden, E., 1961. Type Collections of Drosophilidae. *Beitrage zur Entomologie*, **11**(2): 160.

Beckingham, K. M., J. D. Armstrong, M. J. Texada, R. Munjaal & D. A. Baker, 2005. *Drosophila melanogaster* - the model organism of choice for the complex biology of multi-cellular organisms. *Gravitational and Space Biology Bulletin,* **18**(2): 17-29.

Beintema, J. J., W. T. Stam, B. Hazes & M. P. Smidt, 1994. Evolution of arthropod hemocyanins and insect storage proteins (hexamerins). *Molecular Biology and Evolution,* **11**(3): 493-503.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell & E. W. Sayers, 2013. GenBank. *Nucleic Acids Research,* **41**(Database issue): D36-42.

Bergsten, J., 2005. A review of long-branch attraction. *Cladistics,* **21**(2): 163-193.

Bininda-Emonds, O. R. P., 2005. Supertree Construction in the Genomic Age. *Methods in Enzymology,* **395**: 745-757.

Bocak, L., C. Barton, A. Crampton‐Platt, D. Chesters, D. Ahrens & A. P. Vogler, 2014. Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Systematic Entomology,* **39**(1): 97-110.

Bock, I. R., 1982. Drosophilidae of Australia. V. Remaining genera and synopsis (Insecta: Diptera). *Australian Journal of Zoology,* **30**(89): 1-164.

Bogart, K. & J. Andrews, 2006. Extraction of total RNA from Drosophila. *Center for Genomics and Bioinformatics CGB Technical Report 10.*

Bosch, J., 1992. Parasitism in wild and managed populations of the almond pollinator *Osmia cornuta* Latr. (Hymenoptera: Megachilidae). *Journal of Apicultural Research,* **31**(2): 77-82.

Brake, I. & G. Bächli, 2008. Drosophilidae (Diptera). In I. Brake & G. Bächli (eds.), *World Catalogue of Insects. Vol. 9.* Apollo Books: Stenstrup.

Brower, A. V., 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences,* **91**(14), 6491-6495.

Burman, J. L., L. S. Itsara, E. B. Kayser, W. Suthammarak, A. M. Wang, M. Kaeberlein, M. M. Sendensky & L. J. Pallanck, 2014. A *Drosophila* model of mitochondrial disease caused by a complex I mutation that uncouples proton pumping from electron transfer. *Disease models & mechanisms,* **7**(10): 1165-1174.

Chang, J. L., H. V. Lin, T. A. Blauwkamp & K. M. Cadigan, 2008. Spenito and Split ends act redundantly to promote Wingless signaling. *Developmental Biology,* **314**(1): 100-111.

Chen , H.W., 2000. A phylogenetic and taxonomic study on the subfamily Steganinae (Diptera , Drosophilidae). PhD Thesis. Graduate School of Environmental Earth Science, Hokkaido University, Japan.

Chen, H.-W., M. Toda & B.-C. Wang, 2005. A taxonomic revision of the genus *Pseudostegana* Okada, 1978 (Diptera, Drosophilidae). *Insect Systematics & Evolution,* **36**(4): 407-442.

Clausen, C. P. & P. A. Berry, 1932. *The citrus blackfly in Asia, and the importation of its natural enemies into tropical America.* US Department of Agriculture.

Colless, D. H. & D. K. McAlpine, 1991. Diptera (Flies). In C. S. I. R. S. O. (ed.), *Insects of Australia. A Textbook for Students and Research Workers* Melbourne University Press. Carlton South, Victoria. Vol. 2, Pp. 717-786.

Collin, J. E., 1902. Note on *Acletoxenus syrphoides*, Frauenfeld. *Entomologist's Monthly Magazine, 2nd Series,* **38**: 1-3.

Cornman, R. S., 2010. The distribution of GYR- and YLP-like motifs in *Drosophila* suggests a general role in cuticle assembly and other protein-protein interactions. *PLoS One,* **5**(9).

Courtney, G., B. Sinclair & R. Meier, 2000. Morphology and terminology of Diptera larvae. *Contributions to a Manual of Palaearctic Diptera,* **1**: 85-161.

Csikós, G., M. Lippai, T. Lukácsovich, G. Juhász, L. Henn, M. Erdélyi, P. Maróy & M. Sass, 2009. A novel role for the *Drosophila* epsin (lqf): involvement in autophagy. *Autophagy,* **5**(5): 636-648.

Culik, M. P. & J. A. Ventura, 2009. New species of *Rhinoleucophenga,* a potential predator of pineapple mealybugs. *Pesquisa Agropecuária Brasileira,* **44**(4): 417-420.

De Queiroz, A. & J. Gatesy, 2007. The supermatrix approach to systematics. *Trends in Ecology & Evolution,* **22**(1): 34-41.

Desalle, R., 1992. The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. *Molecular Biology and Evolution,* **9**(5): 905-916.

Desalle, R. & D. A. Grimaldi, 1991. Morphological and Molecular Systematics of the Drosophilidae. *Annual Review of Ecology and Systematics,* **22**: 447-475.

Dos Santos, G., A. J. Schroeder, J. L. Goodman, V. B. Strelets, M. A. Crosby, J. Thurmond, D. B. Emmert & W. M. Gelbart, 2015. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic acids research,* **43**(D1): D690-D697.

Dowding, V. M., 1967. The function and ecological significance of the pharyngeal ridges occurring in the larvae of some cyclorrhaphous Diptera. *Parasitology*, *57*(02), 371-388.

Duda, O., 1924. Beitrag zur Systematik der Drosophiliden unter besonderer Berucksichtigung der palaarktischen u. orientalischen Arten (Dipteren). *Archiv für Naturgeschichte,* **90** (3): 172-234.

Duda, O., 1936. XXXI.—Weitere neue afrikanische und orientalische akalyptrate Musciden (Dipt.) des British Museum. *The Annals and Magazine of Natural History*, **18**(105): 337-351.

Elvin, C. M., T. Vuocolo, R. D. Pearson, I. J. East, G. A. Riding, C. H. Eisemann & R. L. Tellam, 1996. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*. cDNA and deduced amino acid sequences. *The Journal of biological chemistry,* **271**(15): 8925-8935.

Elzinga, R. J. & A. B. Broce, 1986. Labellar modifications of Muscomorpha flies (Diptera). *Annals of the Entomological Society of America,* **79**(1): 150-209.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution,* **17**(6): 368-376.

Fisher, B., R. Weiszmann, E. Frise, A. Hammonds, P. Tomancak, A. Beaton, B.Berman, E. Quan, S. Shu, S. Lewis, G. Rubin, C. Barale, E. Laguertas, J. Quinn, A. Ghosh, V. Hartenstein, M. Ashburner & S. Celniker, 2012. *Berkeley Drosophila Genome Project.* Updated on 1 July 2015. Retrieved on 12 April 2015 from http://www.fruitfly.org/cgi-bin/ex/insitu.pl.

FlyBase Curators, Swiss-Prot Project Members, InterPro Project Members, 2004. *Gene Ontology annotation in FlyBase through association of InterPro records with GO terms.* Available from http://flybase.org/.

Frauenfeld, G. R. V., 1868. Zoologische Miscellen. XIV & XV. *Verhandlungen der Zoologisch-Botanischen Gesellschaft in Wien,* **18**: 1885-1899.

Gao, J.-J., Y.-G. Hu, M. J. Toda, T. Katoh & K. Tamura, 2011. Phylogenetic relationships between *Sophophora* and *Lordiphosa*, with proposition of a hypothesis on the vicariant divergences of tropical lineages between the Old and New Worlds in the family Drosophilidae. *Molecular phylogenetics and evolution,* **60**(1): 98-107.

Goldstein, L. S. B. & S. Gunawardena, 2000. Table S2. *Identified cytoskeletal protein coding genes in Drosophila*, **150**.

Goloboff, P. A., J. S. Farris & K. C. Nixon, 2008. TNT, a free program for phylogenetic analysis. *Cladistics,* **24**(5): 774-786.

Greener, M. J. & R. G. Roberts, 2000. *Conservation of components of the dystrophin complex in Drosophila*, **482**: 13-18.

Grimaldi, D., 1994. Description and immature stages of *Hirtodrosophila batracida* sp. n.(Diptera: Drosophilidae), a predator of frog embryos. *Insect Systematics & Evolution,* **25**(2): 129-136.

Grimaldi, D. A., 1990. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bulletin of the American Museum of Natural History,* **197**: 1-139.

Gupta, A., V. Naveen & C. Malathi, 2013. *Indian fauna of Pteromalidae*. NPIB Project, NBAIR (ICAR), India. Updated on April 2015. Retrieved on 10 January 2015 from http://www.nbair.res.in/Pteromalidae/index.php.

G. Hervé, 1989. *Allosteric Enzymes*, p. 204, Boca Raton: CRC Press.

Hershko, A., 1991. The ubiquitin pathway for protein degradation. *Trends in biochemical sciences*, **16**(7): 265-268.

Hodges, G. S. & G. A. Evans, 2005. An identification guide to the whiteflies (Hemiptera: Aleyrodidae) of the Southeastern United States. *Florida Entomologist,* **88**(4): 518-534.

Hollfelder, D., M. Frasch & I. Reim, 2014. Distinct functions of the laminin β LN domain and collagen IV during cardiac extracellular matrix formation and stabilization of alary muscle attachments revealed by EMS mutagenesis in Drosophila. *BMC Developmental Biology,* **14**: 26.

Hu, Y. G. & M. J. Toda, 2001. Polyphyly of *Lordiphosa* and its relationships in Drosophilinae (Diptera: Drosophilidae). *Systematic Entomology,* **26**(1): 15-31.

Jänsch, L., V. Kruft, U. K. Schmitz & H. P. Braun, 1998. Unique composition of the preprotein translocase of the outer mitochondrial membrane from plants. *The Journal of biological chemistry*, **273**(27): 17251-17257. (eng)

Katoh, K. & D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution,* **30**(4): 772-780.

Kelemen-Valkony, I., M. Kiss, J. Csiha, A. Kiss, U. Bircher, J. Szidonya, P. Maróy, G. Juhász, O. Komonyi, K. Csiszár & M. Mink, 2012. *Drosophila* basement membrane collagen col4a1 mutations cause severe myopathy. *Matrix biology. Journal of the International Society for Matrix Biology,* **31**(1): 29-37.

King-Jones, K., J. P. Charles, G. Lam & C. S. Thummel, 2005. The ecdysone-induced DHR4 orphan nuclear receptor coordinates growth and maturation in *Drosophila*. *Cell,* **121**(5): 773-784.

Kwiatowski, J., M. Krawczyk, M. Jaworski, D. Skarecky & F. J. Ayala, 1997. Erratic evolution of glycerol-3-phosphate dehydrogenase in *Drosophila*, *Chymomyza*, and *Ceratitis*. *Journal of Molecular Evolution,* **44**(1): 9-22.

Kwiatowski, J., D. Skarecky, K. Bailey & F. J. Ayala, 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn Sod gene. *Journal of Molecular Evolution,* **38**(5): 443-454.

Lambkin, T. A. & M. P. Zalucki, 2010. Long-term efficacy of *Encarsia dispersa* Polaszek (Hymenoptera: Aphelinidae) for the biological control of *Aleurodicus dispersus* Russell (Hemiptera: Aleyrodidae) in tropical monsoon Australia. *Australian Journal of Entomology,* **49**(2): 190-198.

Lanio, W., U. Swida & H. Kress, 1994. Molecular cloning of *the Drosophila virilis* larval glue protein gene Lgp-3 and its comparative analysis with other Drosophila glue protein genes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* **1219**(2): 576-580.

Levinson, H. Z., 1976. The defensive role of alkaloids in insects and plants. *Experientia,* **32**(4): 408-411.

Li, L., C. J. Stoeckert & D. S. Roos, 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research,* **13**(9): 2178-2189.

Loew, H., 1864. *Gitona formosa*: eine neue deutsche Art. *Wiener Entomologische Monatsschrift,* **8**: 366-368.

Mace, K. & A. Tugores, 2004. The product of the split ends gene is required for the maintenance of positional information during *Drosophila* development. *BMC Developmental Biology,* **4**(1): 15.

Maddison, W. & D. Maddison, 2015. Mesquite: a modular system for evolutionary analysis. Version 3.03. Available from http://mesquiteproject.org.

Malloch, J. R., 1929. LXVI. Exotic Muscaridae (Diptera). XXV. *Journal of Natural History*, **3**(17): 545-564.

Maresh, J. J., L.-A. Giddings, A. Friedrich, E. A. Loris, S. Panjikar, B. L. Trout, J. Stöckigt, B. Peters & S. E. O'connor, 2008. Strictosidine synthase: mechanism of a Pictet-Spengler catalyzing enzyme. *Journal of the American Chemical Society*, **130**(2): 710-723.

Markow, T. A. & P. M. O'grady, 2006. Chapter 1 - Phylogenetic relationships of Drosophilidae. In T. a. M. M. O'grady (ed.), *Drosophila* Academic Press: San Diego. Pp. 3-64.

Martin, J., 1987. An identification guide to common whitefly pest species of the world (Homopt Aleyrodidae). *International Journal of Pest Management*, **33**(4): 298-322.

McAlpine, J., 1968. An annotated key to drosophilid genera with bare or micropubescent aristae and a revision of *Paracacoxenus* (Diptera: Drosophilidae). *The Canadian Entomologist*, **100**(05): 514-532.

McGregor, R. & D. Gillespie, 2004. Olfactory responses of the omnivorous generalist predator *Dicyphus hesperus* to plant and prey odours. *Entomologia experimentalis et applicata*, **112**(3): 201-205.

McMahon, M. M. & M. J. Sanderson, 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Systematic biology*, **55**(5): 818-836.

Meier, R., K. Shiyang, G. Vaidya & P. K. Ng, 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic biology*, **55**(5): 715-728.

Mentzelos, I. A., 1967. Contribution to the study of the entomophagous insects of *Siphoninus phillyreae* Halid. (=inaequalis Gautier) (Aleyrodidae) on pear treas in central Macedonia. *Report of the Plant Protection and Agricultural Research Station, Thessaloniki*, **3**: 92-102.

Miller, M., W. Pfeiffer & T. Schwartz, 2010. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* Paper presented at the Gateway Computing Environments Workshop (GCE), 2010.

Mitchell, A., H.-Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, Maxim, C. Rato, S. Yong, A. Bateman, M. Punta, T. K. Attwood, C. J. A. Sigrist, N. Redaschi, C. Rivoire, I. Xenarios, D. Kahn, D. Guyot, P. Bork, I. Letunic, J. Gough, M. Oates, D. Haft, H. Huang, D. A. Natale, C. H. Wu, C. Orengo, I. Sillitoe, H. Mi, P. D. Thomas & R. D. Finn, 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, **43**(D1): D213-D221.

Morrison, D. A. & J. T. Ellis, 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of *Apicomplexa*. *Molecular Biology and Evolution*, **14**(4): 428-441.

Nomikou, M., R. Meng, R. Schraag, M. W. Sabelis & A. Janssen, 2005. How predatory mites find plants with whitefly prey. *Experimental & applied acarology*, **36**(4): 263-275.

Noyes, J., 2015. *Universal Chalcidoidea Database.* Updated on April 2015. Retrieved on 10 January 2015 from http://www.nhm.ac.uk/chalcidoids.

O'grady, P. & R. Desalle, 2008. Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). *Biology Letters*, **4**(2): 195-199.

O'connor, T. K., P. T. Humphrey, R. T. Lapoint, N. K. Whiteman & P. M. O'grady, 2014. Microbial interactions and the ecology and evolution of Hawaiian Drosophilidae. *Frontiers in microbiology*, **5**.

Ogden, T. H. & M. S. Rosenberg, 2006. Multiple sequence alignment accuracy and phylogenetic inference. *Systematic Biology*, **55**(2): 314-328.

Okada, T., 1977. Family Drosophilidae. *A Catalog of the Diptera of the Oriental Region*, **3**: 342-387.

Okada, T., 1989. A Proposal of Establishing Tribes for the Family Drosophilidae with Key to Tribes and Genera (Diptera) : Taxonomy and Systematics. *Zoological Science*, **6**.

Otranto, D., J. R. Stevens, G. Testini, C. Cantacessi & J. Máca, 2008. Molecular characterization and phylogenesis of Steganinae (Diptera, Drosophilidae) inferred by the mitochondrial cytochrome c oxidase subunit 1. *Medical and Veterinary Entomology*, **22**(1): 37-47.

Overstreet, E., X. Chen, B. Wendland & J. A. Fischer, 2003. Either part of a *Drosophila* Epsin protein, divided after the ENTH Domain, functions in endocytosis of Delta in the developing eye. *Current Biology*, **13**(10): 854-860.

Parchami-Araghi, M. & S. Farrokhi, 1995. *Acletoxenus formosus* Loew (Dip: Drosophilidae), predator of immature stages of whiteflies in Iran. *Journal of Entomological Society of Iran*, **15**.

Pélandakis, M. & M. Solignac, 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. *Journal of molecular evolution*, **37**(5): 525-543.

Pelov, V. & G. Trenchev, 1973. (*Siphonius phillyrea* Hal. And its entomophages). *Rastitelna Zashchita*, **21**(11): 26-27.

Pestnet, 2011. *Aleurotrachelus trachoides, biocontrol, FSM*. PestNet. Updated on 2015. Retrieved on 10 April. 2015 from http://www.pestnet.org/SummariesofMessages/Crops/Vegetables/Capsicumchilli/Aleurotrachelustrachoides,biocontrol,FSM.aspx.

Peyrefitte, S., D. Kahn & M. Haenlin, 2001. New members of the *Drosophila* Myc transcription factor subfamily revealed by a genome-wide examination for basic helix-loop-helix genes. *Mechanisms of Development*, **104**(1-2): 99-104.

Pol, D. & I. H. Escapa, 2009. Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics*, **25**(5): 515-527.

Rebers, J. E. & J. H. Willis, 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem Molecular Biology*, **31**(11): 1083-1093.

Remsen, J. & P. O'grady, 2002. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Molecular phylogenetics and evolution*, **24**(2): 249-264.

Roberts, M. J., 1971. The structure of the mouthparts of some galypterate dipteran larvae in relation to their feeding habits. *Acta zoologica*, *52*(2), 171-188.

Romero-Calderon, R., R. M. Shome, A. F. Simon, R. W. Daniels, A. Diantonio & D. E. Krantz, 2007. A screen for neurotransmitter transporters expressed in the visual system of *Drosophila melanogaster* identifies three novel genes. *Developmental Neurobiology*, **67**(5): 550-569.

Roos, J., P. J. Digregorio, A. V. Yeromin, K. Ohlsen, M. Lioudyno, S. Zhang, O. Safrina, J. A. Kozak, S. L. Wagner, M. D. Cahalan, G. Velicelebi & K. A. Stauderman, 2005. STIM1, an essential and conserved component of store-operated Ca^{2+} channel function. *Journal of Cell Biology*, **169**(3): 435-445.

Russo, C., N. Takezaki & M. Nei, 1995. Molecular phylogeny and divergence times of drosophilid species. *Molecular biology and evolution,* **12**(3): 391-404.

Sanderson, M. J., A. Purvis & C. Henze, 1998. Phylogenetic supertrees: assembling the trees of life. *Trends in Ecology & Evolution,* **13**(3): 105-109.

Schmitz, H. J., 2010. *Genética, ecologia e evolução de drosofilídeos (Insecta, Diptera) associados a flores.* Ph.D. Thesis, Universidade Federal do Rio Grande do Sul, Brazil.

Schloss, P. D., & J. Handelsman, 2006. Toward a census of bacteria in soil. *PLoS Computational Biology,* **2**(7): e92.

Schulz, S., 1998. Insect-Plant Interactions - Metabolism of Plant Compounds to Pheromones and Allomones by Lepidoptera and Leaf Beetles. *European journal of organic chemistry,* **1998**(1): 13-20.

Séguy, E., 1951. Dipteres mineurs de Madagascar. *Mémoires de l'Institut Scientifique de Madagascar (A),* **5**: 309-321.

Seidelmann, K., 1999. The function of the vestibulum in nests of a solitary stem-nesting bee, *Osmia rufa* (L.). *Apidologie* **30** (1): 19-29.

Seto, Y. & K. Tamura, 2013. Extensive Differences in Antifungal Immune Response in Two Drosophila Species Revealed by Comparative Transcriptome Analysis. *International Journal of Genomics,* **2013**: 542139.

Sidorenko, V., 2002. Phylogeny of the tribe Steganini Hendel and some related taxa (Diptera, Drosophilidae). *Far Eastern Entomologist,* **111**: 1-20.

Simmons, M. P. & P. A. Goloboff, 2014. Dubious resolution and support from published sparse supermatrices: The importance of thorough tree searches. *Molecular phylogenetics and evolution,* **78**: 334-348.

Siu-Ting, K., D. Pisani, C. J. Creevey & M. Wilkinson, 2015. Concatabominations: Identifying unstable taxa in morphological phylogenetics using a heuristic extension to safe taxonomic reduction. *Systematic Biology,* **64**(1): 137-143.

Smythe, A. B., M. J. Sanderson & S. A. Nadler, 2006. Nematode small subunit phylogeny correlates with alignment parameters. *Systematic biology,* **55**(6): 972-992.

Spencer, W. P., 1938. *Drosophila virilis americana*, a new subspecies. *Genetics,* **23**: 169.

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, **30** 1312 – 1313

Stec, I., S. B. Nagl, G. J. Van Ommen & J. T. Den Dunnen, 2000. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS letters*, **473**(1): 1-5.

Stogios, P. J., G. S. Downs, J. J. S. Jauhal, S. K. Nandra & G. G. Privé, 2005. Sequence and structural analysis of BTB domain proteins. *Genome Biology*, **6**(10): R82.

Su, Y., J. Lu & H. Chen, 2013. The genus *Leucophenga* (Diptera, Drosophilidae), part I: the abbreviata species group from the Oriental region with morphological and molecular evidence. *Zootaxa*, **3637**(3): 361-373.

Sukumaran, J. & M. T. Holder, 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, **26**(12): 1569-1571.

Sultana, F., Y.-G. Hu, M. J. Toda, K. Takenaka & M. Yafuso, 2006. Phylogeny and classification of *Colocasiomyia* (Diptera, Drosophilidae), and its evolution of pollination mutualism with aroid plants. *Systematic Entomology*, **31**(4): 684-702.

Sureshan, P. & T. Narendran, 2000. Pteromalidae (Chalcidoidea: Hymenoptera) from India with The Description of a New Species. *Entomon*, **25**(2): 117-128.

Syed, Z. A., T. Härd, A. Uv & I. F. Van Dijk-Härd, 2008. A potential role for *Drosophila* mucins in development and physiology. *PLoS ONE*, **3**(8): e3041.

Talavera, G. & J. Castresana, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*, **56**(4): 564-577.

Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**(12): 2725-2729.

Tarrío, R., F. Rodríguez-Trelles & F. J. Ayala, 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Molecular Biology and Evolution*, **18**(8): 1464-1473.

Tatarenkov, A., M. ŽUrovcová & F. J. Ayala, 2001. Ddc and amd sequences resolve phylogenetic relationships of *Drosophila*. *Molecular Phylogenetics and Evolution*, **20**(2): 321-325.

Theuniprotconsortium, 2015. UniProt: a hub for protein information. *Nucleic Acids Research*, **43**(D1): D204-D212.

Throckmorton, L., 1962. The problem of phylogeny in the genus *Drosophila*. *Univ Texas Publ*, **6205**: 207-343.

Throckmorton, L. H., 1975. The phylogeny, ecology and geography of *Drosophila*. *Handbook of genetics*, **3**: 421-469.

Tsacas, L. & R. Disney, 1974. Two new African species of *Drosophila* (Diptera, Drosophilidae) whose larvae feed on *Simulium* larvae (Diptera, Simuliidae). *Zeitschrift fur Tropenmedizin und Parasitologie*, **25**(3):360-77.

Vaidya, G., D. J. Lohman & R. Meier, 2011. SequenceMatrix: concatenation software for the fast assembly of multi‐gene datasets with character set and codon information. *Cladistics*, **27**(2): 171-180.

Van Der Linde, K., D. Houle, G. S. Spicer & S. J. Steppan, 2010. A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genetics research*, **92**(01): 25-38.

Vayssière, P., 1953. Rapport de la Commission pour les recherches sur la lutte biologique contre les ennemis des cultures. *Comptes Rendus de l'Assemblée générale de l'Union Internationale des Sciences Biologique, Nice*: 12.

Vogt, R. G., G. D. Prestwich & M. R. Lerner, 1991. Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *Journal of neurobiology*, **22**(1): 74-84.

Walker, K., 2008. *Aleurotrachelus whitefly (Aleurotrachelus trachoides)* PaDIL webiste. Updated on 25 November 2011. Retrieved on 14 January 2015 from http://www.padil.gov.au.

Wheeler, M. R., 1952. The Drosophilidae of the Nearctic region, exclusive of the genus *Drosophila*. *University of Texas Publications*, **5204**: 162-218.

Wheeler, M. R. & H. Takada, 1971. Male genitalia of some representative genera of American Drosophilidae. *University of Texas Publications*,**7103**: 225-240.

Wilkinson, M., 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology*, **44**(4): 501-514.

Wilkinson, M., 2003. Missing entries and multiple trees: instability, relationships, and support in parsimony analysis. *Journal of Vertebrate Paleontology*, **23**(2): 311-323.

Williams, C., 1923. A froghopper damaging cacao in Panama. *Bulletin of entomological Research,* **13**(03): 271-274.

Windler, S. L. & D. Bilder, 2010. Endocytic internalization routes required for delta/notch signaling. *Current Biology,* **20**(6): 538-543.

Wirth, W. W., 1952. Two new spider egg predators from the Hawaiian Islands (Diptera: Drosophilidae). *Proceedings of the Hawaiian Entomological Society,* **14**(3): 415-417.

Wu, M., S. Chatterji & J. A. Eisen, 2012. Accounting for alignment uncertainty in phylogenomics. *PLoS One,* **7**(1): e30288.

Xu, D. & R. Nussinov, 1998. Favorable domain size in proteins. *Folding and Design,* **3**(1): 11-17.

Yang, J. & T. Warnow, 2011. Fast and accurate methods for phylogenomic analyses. *BMC bioinformatics,* **12**(Suppl 9): S4.

Yang, Z. & B. Rannala, 2012. Molecular phylogenetics: principles and practice. *Nat Rev Genet,* **13**(5): 303-314.

Yassin, A., 2013. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the postgenomic era. *Systematic Entomology,* **38**(2): 349-364.

Yassin, A., J.-L. Da Lage, J. R. David, M. Kondo, L. Madi-Ravazzi, S. R. Prigent & M. J. Toda, 2010. Polyphyly of the *Zaprionus* genus group (Diptera: Drosophilidae). *Molecular Phylogenetics and Evolution,* **55**(1): 335-339.

Yu, G., L. Wu, J. Lu & H. Chen, 2012. Discovery of a predaceous drosophilid *Acletoxenus indicus* Malloch in South China, with descriptions of the taxonomic, ecological and molecular characters (Diptera: Drosophilidae). *Journal of Natural History,* **46**(5-6): 349-354.

Appendix

Table A1. *Acletoxenus* cf. *indicus* COI Sequences

ID	Sequence
A. formosus morphotype male 1	gatattggtactctatattttattttcggggcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcat ccaggggcattaattggaaacgatcaaatttataatgttatcgttacagctcatgcatttattatgatttttttttatagttataccaattat aattggaggatttggaaattgattagtacctttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgact tttacctcctgcactttcacttttactaataagaagtatagttgaaaacggagctggtacaggatgaacagtttacccaccttttatct tcaggaattgctcacggaggagcttctgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaat tttattacaactgttattaatatacgatcaactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttat tacttctatctttacctgttttagctggagcaattactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggagga ggagatcctatttttatatcaacatttatttgatttt?? ?????ttgttattagacaagaagctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttatta ggatttattgtttgagctcatcatatatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgc tgttcctacaggaattaaaatttttagttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattagga tttgttttcttatttacagtaggaggattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagt agcccattttcattatgtactatctataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacat taaataataaattattaaaaactcaatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtat acctcgacgatattctgattacccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaatttt attttttttttacattttatgagaaagcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaa tacaccgccagctg
A. formosus morphotype male 2	ctctatattttattttcggggcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggc attaattggaaacgatcaaatttataatgttatcgttacagctcatgcatttattatgatttttttttatagttataccaattataattggag gatttggaaattgattagtacctttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgacttttacctcc tgcactttcacttttactaataagaagtatagttgaaaacggagctggtacaggatgaacagtttacccaccttttatcttcaggaat tgctcacggaggagcttctgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattaca actgttattaatatacgatcaactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctat ctttacctgttttagctggagcaattactatattattaactgatcgaaattttaatacatcatttttgatcctgca?????????????? ?? ??????tattattagacaagaagctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttatta ggatttattgtttgagctcatcatatatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgc tgttcctacaggaattaaaatttttagttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattagga tttgttttcttatttacagtaggaggattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagt agcccattttcattatgtactatctataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacat taaataataaattattaaaaactcaatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtat acctcgacgatattctgattacccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaatttt attttttttttacattttatgagaaagcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaa tacaccgccagc
A. indicus morphotype male 1	tcggggcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaac gatcaaatttataatgttatcgttacagctcatgcatttattatgatttttttttatagttataccaattataattggaggatttggaaattga ttagtacctttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgacttttacctcctgcactttcacttttt actaataagaagtatagttgaaaacggagctggtacaggatgaacagtttacccaccttttatcttcaggaattgctcacggagg agcttctgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatata cgatcaactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctatctttacctgttttag ctggagcaattactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggaggaggagatcctatttttatatcaac a?? ??agctattggtttattaggatttattg tttgagctcatcatatatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctaca ggaattaaaatttttagttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttctt atttacagtaggaggattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccatt ttcattatgtactatctataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataat aaattattaaaaactcaatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcga cgatattctgattacccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaatttttatttttttt tacattttatgagaaag

Table A1. *Acletoxenus* cf. *indicus* COI Sequences (cont'd)

ID	Sequence
A. indicus morphotype male 2	tgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatcaaatt tataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattgattagtacct ttaatattaggagctcctgatatagcatttcctcgaataaataatataagtttttgacttttacctcctgcactttcacttttactaataag aagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagcttctgtt gatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacgatcaact ggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctatctttacctgttttagctggagcaa ttactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggaggaggagatcctatctc???atattattagacaagaag ctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatcata tatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaatttttt agttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtaggag gattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactatc tataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaact caatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattctgattac ccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacattttatgagaa agcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaatacaccgccagctg
A. quadristriatus morphotype male 1	cttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatcaa atttataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattgattagta ccctttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgacttttacctcctgcactttcacttttactaat aagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagcttct gttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacgatca actggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctatctttacctgttttagctggag caattactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggaggaggagatcctatttatatcaaca????? ??tctcatattattagacaagaag ctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatcata tatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaatttttt agttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtaggag gattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactatc tataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaact caatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattctgattac ccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacattttatgagaa agcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaatacaccgcc
A. quadristriatus morphotype male 2	gcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatca aatttataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattgattagt accttttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgacttttacctcctgcactttcacttttactaa taagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagcttc tgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacgatc aactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctatctttacctgttttagctggga gcaattactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggaggaggagatcctatttatatcaacatttat ctc???atattattagacaagaag ctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatcata tatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaatttttt agttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtaggag gattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactatc tataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaact caatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattctgattac ccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacattttatgagaa agcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatga
A. formosus morphotype female 1	gcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatca aatttataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattgattagt accttttaatattaggagctcctgatatagcatttcctcgaataaataatataagttttgacttttacctcctgcactttcacttttactaa taagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagcttc tgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacgatc aactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttattacttctatctttacctgttttagctggga gcaattactatattattaactgatcgaaattttaatacatcatttttgatcctgcaggaggaggagatcctatttatatcaacat??? ??atttctcatattattagacaagaag ctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatcata tatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaatttttt agttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtaggag gattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactatc tataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaact caatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattgtgattac ccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacattttatgagaa agcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaatacaccgccagctgaacata

Table A1. *Acletoxenus* cf. *indicus* COI Sequences (cont'd)

ID	Sequence
A. formosus morphotype female 2	caaatttataatgttatcgttacagctcatgcatttattatgatttttttttatagttataccaattataattggaggatttggaaattgatta gtacctttaatattaggagctcctgatatagcatttcctcgaataaataatataagtttttgacttttacctcctgcactttcacttttact aataagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagc ttctgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacga tcaactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttttattacttctatctttacctgttttagctgg agcaattactatattattaactgatcgaaattttaatacatcatttttttgatcctgcaggaggaggagatcctattttatatcaacattta ttttgatttttttggtcatccagaagtt?? ??????aaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatcatatat ttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaatttttag ttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtaggagga ttaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactatctat aggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaactca attttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattctgattaccc agatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattacta
A. indicus morphotype female 1	tcggggcttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaac gatcaaatttataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattga ttagtacctttaatattaggagctcctgatatagcatttcctcgaataaataatataagtttttgacttttacctcctgcactttcacttttt actaataagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggagg agcttctgttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatata cgatcaactggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttttattacttctatctttacctgttttag ctggagcaattactatattattaactgatcgaaattttaatacatcatttttttgatcctgcaggaggaggagatcctactc??????? ??atattattag acaagaagctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgag ctcatcatatatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaat taaaatttttagttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttac agtaggaggattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcatt atgtactatctataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaatt attaaaaactcaatttttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgat attctgattacccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacat tttatgagaaagcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaatacaccgcc
A. indicus morphotype female 2	cttgatctggaatagtaggaacttctttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatcaa atttataatgttatcgttacagctcatgcatttattatgatttttttatagttataccaattataattggaggatttggaaattgattagta ccttttaatattaggagctcctgatatagcatttcctcgaataaataatataagtttttgacttttacctcctgcactttcacttttactaat aagaagtatagttgaaaacggagctggtacaggatgaacagtttacccacctttatcttcaggaattgctcacggaggagcttct gttgatttagctatttttttcattacatttagcgggaatttcttcaattttaggtgcagtaaattttattacaactgttattaatatacgatca actggaattactttagatcgtatacctttatttgtttgatcagttgtaattactgccttttttattacttctatctttacctgttttagctggag caattactatattattaactgatcgaaattttaatacatcatttttttgatcct?????????????????????????????????????? ??tattattagacaaga agctggaaaaaaagaaacatttggatcattaggtataatctatgctatattagctattggtttattaggatttattgtttgagctcatca tatatttacagtaggaatagacgtagatacacgagcatattttacatcagctacaataattattgctgttcctacaggaattaaaattt ttagttgattagcaactttacatggaactcaattaacttattctcctgcaattttatgaacattaggatttgttttcttatttacagtagga ggattaacaggagtagtattagctaattcatctattgacattgtacttcatgatacatattatgtagtagcccattttcattatgtactat ctataggtgcagtatttgctattatagcaggatttattcattgatatccattatttacaggattaacattaaataataaattattaaaaac tcaattttaattatatttatcggagtaaatttaacttttttcctcaacatttttaggattagcaggtatacctcgacgatattctgatta cccagatgcttatacaacatgaaatattatttcaacaattgggtcttcaatttcattactaggaattttatttttttttacattttatgaga aagcttaatttctcaacgtaaggtaatttatccaattcaattaaattcatcaattgaatgatttcaaaatacaccgccagctg

Table A2. Whitefly prey COI Sequences

ID	Sequence
1	CACTGTCTATAAATTCAGCACACAGGGGCTGTTCAGTTGATATATCTATTTTATCT TTACATGTTGCCGGGGCTTCATCTATTTTGGGGGCAATAAATTTTATTGTAACTAT TTTTAATATGCGAGTTTTGGGGGTTAGTTTGGATTTAATAAGTTTGTTTGTTTGATC TGTTTTAATTACTGTCTTTTTACTGTTAATTTCGTTGCCCGTTTTGGCCGGGGCTAT TACAATACTTCTTTTTGATCGTAATTTTAACAGGTCTTTTTATGACCCCCTCGGGG GCGGCGATCCCATTTTGTATCAACATTTGTTTT
2	CACTGTCTATAAATTCAGCACACAGGGGCTGTTCAGTTGATATATCTATTTTATCT TTACATGTTGCCGGGGCTTCATCTATTTTGGGGGCAATAAATTTTATTGTAACTAT TTTTAATATGCGAGTTTTGGGGGTTAGTTTGGATTTAATAAGTTTGTTTGTTTGATC TGTTTTAATTACTGTCTTTTTACTGTTAATTTCGTTGCCCGTTTTGGCCGGGGCTAT TACAATACTTCTTTTTGATCGTAATTTTAACAGGTCTTTTTATGACCCCCTCGGGG GCGGCGATCCCATTTTGTATCAACATTTGTTT

Table A3. Parasitoid COI Sequences

ID	Sequence
1	CCTCTATCATCTAATATCTCTCATGGAGGACCATCAGTAGATTTATCAAT TTTTT CATTACATATTGCAGGAATTTCTTCAATTATAGGATCAATTAATT TTATTTCAAC TATTTTAAATATAAAAATTTTTAAAATTGAAATTATTCCT CTTTTTGCTTGATCTA TATTATTAACTGCTATTTTATTATTATTATCTTT ACCTGTGTTAGCAGGAGCTATT ACTATACTTTTATTTGATCGTAACTTAA ATACATCATTTTTTGATCCAGCAGGAG GAGGAGATCCAATTTTATATCAA CATTTATTTTGATTTTTTGGCCA
2	ATAATTTTTTTTTTTGTTATACCCGTAATAATAGGAGGATTTGGTAATTA TTTAA TTCCATTAATTTTAGGATCTCCTGATATAGCATTCCCTCGAATAA ATAATATAAG ATTTTGATTATTACCTCCAAGATTAATAATTTTAATTTCT AGGATATTTGTTGGT ACTGGAACAGGTACAGGATGAACAGTTTATCCGCC TCTATCATCTAATATCTCT CATGGAGGACCATCAGTAGATTTATCAATTT TTTCATTACATATTGCAGGAATTT CTTCAATTATAGGATCAATTAATTTT ATTTCAACTATTTTAAATATAAAAATTTT TAAAATTGAAATTATTCCTCT TTTTGCTTGATCTATATTATTAACTGCTATTTTAT TATTATTATCTTTAC CTGTGTTAGCAGGAGCTATTACTATACTTTTATTTGATCGT AACTTAAAT ACATCATTTTTTGATCCAGCAGGAGGAGGA
3	ATTGTTACTACTCATGCTTTCGTTATAATTTTTTTTTTTGTTATACCCGT AATAAT AGGAGGATTTGGTAATTATTTAATTCCATTAATTTTAGGATCTC CTGATATAGCA TTCCCTCGAATAAATAATATAAGATTTTGATTATTACCT CCAAGATTAATAATTT TAATTTCTAGGATATTTGTTGGTACTGGAACAGG TACAGGATGAACAGTTTATC CGCCTCTATCATCTAATATCTCTCATGGAG GACCATCAGTAGATTTATCAATTTT TTCATTACATATTGCAGGAATTTCT TCAATTATAGGATCAATTAATTTTATTTCA ACTATTTTAAATATAAAAAT TTTTAAAATTGAAATTATTCCTCTTTTTGCTTGATC TATATTATTAACTG CTATTTTATTATTATTATCTTTACCTGTGTTAGCAGGAGCTA TTACTATA CTTTTATTTGATCGTAACTTAAATACATCATTTTTTGATCCAGCAGG AGG AGGAGATCCAATTTTATATCAACATTTATTTTGATTT
4	CATCTAATATCTCTCATGGAGGACCATCAGTAGATTTATCAATTTTTTCA TTACA TATTGCAGGAATTTCTTCAATTATAGGATCAATTAATTTTATTTC AACTATTTTA AATATAAAAATTTTTAAAATTGAAATTATTCCTCTTTTTG CTTGATCTATATTATT AACTGCTATTTTATTATTATTATCTTTACCTGTG TTAGCAGGAGCTATTACTATA CTTTTATTTGATCGTAACTTAAATACATC ATTTTTTGATCCAGCAGGAGGAGGAG ATCCA

Table A4. Contigs blasted to species outside of Drosophilidae

Species name	Common name	Species name	Common name
Dictyostelium discoideum	Amoeba	Physcomitrella patens	Moss
Acromyrmex echinatior	Ant	Adoxophyes orana	Moth
Camponotus floridanus	Ant	Cnaphalocrocis medinalis	Moth
Leptanilloides mckennae	Ant	Mus musculus	Mouse
Vollenhovia emeryi	Ant	Angiostrongylus cantonensis	Nematode
Acyrthosiphon pisum	Aphid	Brugia pahangi	Nematode
Burkholderia oklahomensis	Bacteria	Onchocerca ochengi	Nematode
Coraliomargarita akajimensis	Bacteria	Thelazia callipaeda	Nematode
Pantoea ananatis	Bacteria	Toxocara canis	Nematode
Phenylobacterium zucineum	Bacteria	Enterobius vermicularis	Pinworm
Pseudomonas plecoglossicida	Bacteria	Syphacia muris	Pinworm
Roseiflexus	Bacteria	Plasmodium chabaudi	Protozoa
Cerotoma trifurcata	Bettle	Plasmodium falciparum	Protozoa
Bombus terrestris	Bumblebee	Toxoplasma gondii	Protozoan
Cyprinus carpio	Common carp	Galdieria sulphuraria	Red alga
Sphyracephala europaea	Diptera: diopsidae	Tribolium castaneum	Red flour beetle
Echinostoma caproni	Echinoderm	Strongyloides stercoralis	Roundworm
Cynoglossus semilaevis	Fish	Aplysia californica	Sea slug
Protopolystoma xenopodis	Flatworm	Bombyx mori	Silkworm
Schistocephalus solidus	Flatworm	Capsaspora owczarzaki	Cellular eukaryote
Schistosoma margrebowiei	Flatworm	Coturnix coturni	Snail
Schistosoma margrebowiei	Flatworm	Diphyllobothrium latum	Tapeworm
Xenopus laevis	Frog	Solanum lycopersicum	Tomato
Podospora anserina	Fungi	Solanum pennellii	Tomato
Botrytis cinerea	Fungus	Larix kaempferi	Tree (Japanese larch)
Marssonina brunnea	Fungus	Meleagris gallopavo	Turkey
Trichophyton rubrum	Fungus	Athalia rosae	Turnip sawfly
Vitis vinifera	Grape vine	Cotesia sesamiae	Wasp
Brachypodium distachyon	Grass	Microplitis demolitor	Wasp
Volvox carteri	Green algae	Orussus abietinus	Wasp
Natrialba magadii	Halobacteria	Triticum aestivum	Wheat
Sinentomon erythranum	Hexapoda-protura	Bemisia tabaci	Whitefly
Megachile rotundata	Leafcutter bee	Saccharomyces cerevisiae	Yeast
Helobdella robusta	Leech	Danio rerio	Zebrafish

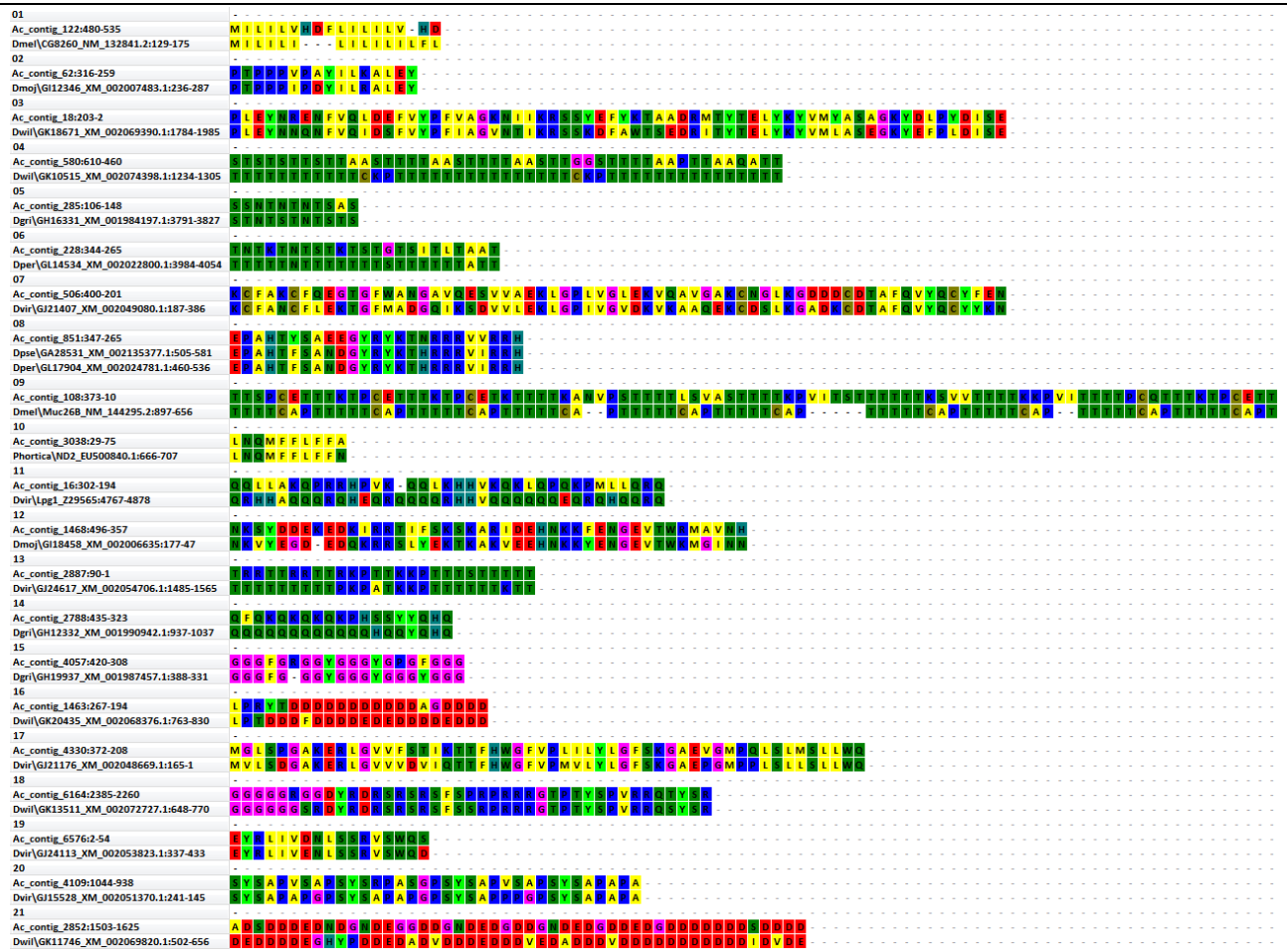

Fig. A1. Alignment of highly expressed *Acletoxenus* sp. contigs and orthologous blast hit

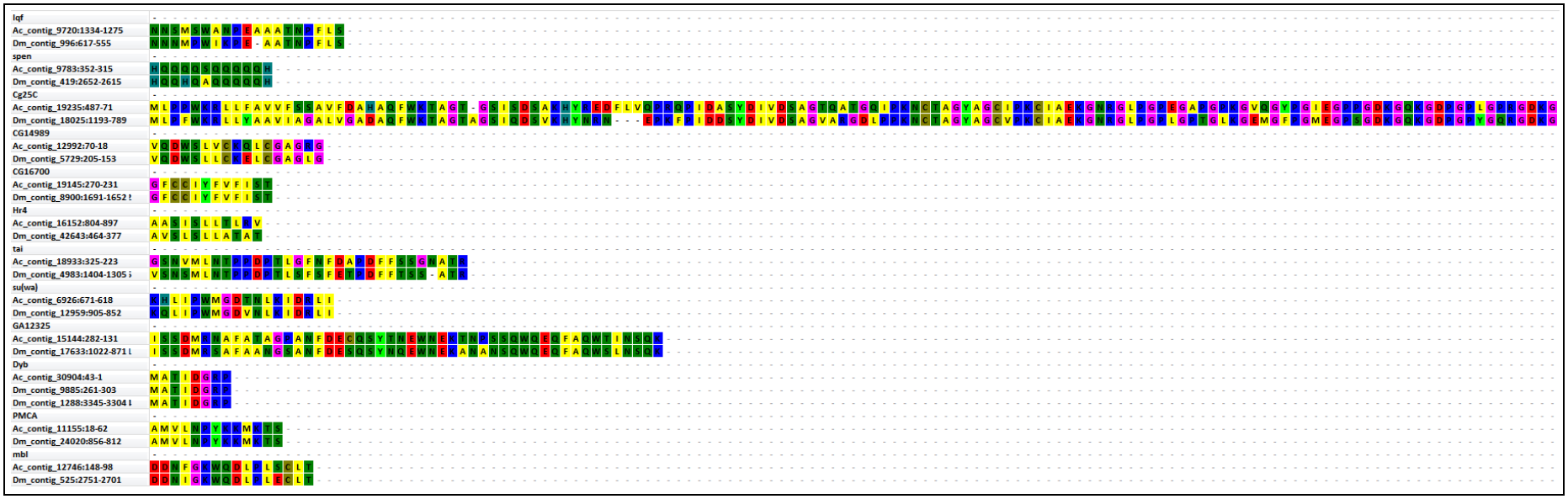

Fig. A2. Alignment of orthologous *Acletoxenus* sp. and *Drosophila melanogaster* contigs

Table A5. List of filter names for gene extraction

Gene	Search term
COI	Co1, CO1, COI, CoI, cox1, COX1, cytochrome c oxidase subunit 1, cytochrome c oxidase subunit I, cytochrome oxidase I, cytochrome oxidase I subunit, cytochrome oxidase subunit, cytochrome oxidase subunit 1, cytochrome oxidase subunit I, cytochrome oxydase subunit I
COII	CO II, co2, COII, CoII, coii, CO-II, cox2, COX2, COXII, cytochrome c oxidase II, cytochrome c oxidase subunit II, cytochrome oxidase II, cytochrome oxidase subunit 2, cytochrome oxidase subunit II, cytochrome oxydase II
COIII	COIII, cox3, COX3, cytochrome c oxidase subunit III, cytochrome oxidase III, cytochrome oxidase subunit 3, cytochrome oxidase subunit III, CO3
12S rRNA	12S ribosomal RNA, s-rRNA, small subunit ribosomal RNA, small ribosomal RNA, 12S small ribosomal RNA, 12S rRNA, 12S rRNA, 12S small ribosomal RNA, 12S small subunit ribosomal RNA, srRNA
16S rRNA	16S ribosomal RNA, l-rRNA, large ribosomal RNA, large subunit ribosomal RNA, 16S rRNA
28SrRNA	28S large subunit ribosomal RNA, 28S ribosomal RNA, 28S ribosomal RNA, D2 domain, 28S rRNA
ADH	Adh, adh, ADH, Adh1, Adh-1, ADH-1, adh-1, Adh2, ADH-2, Adh-2, adh-2, alchohol dehydrogenase, alcohol dehydrogenase, alcohol dehydrogenase 1, alcohol dehydrogenase 2, alcohol dehydrogenase-2, alcohol dehydrpgenase, Adh-Finnegan, Adh-P, Adh-PA, Adh gene, Adh-Twain
Amyrel	amylase related protein, amylase-related protein, Amyrel, amyrel, AMYREL, Amyrel-PA, putative amylase-related protein, putative amylase-related protein AMYREL, alpha amylase-related-protein
per	per, period, period clock protein, period locus protein, period protein, putative period, per protein
Yp1	Yolk protein 1, yolk protein 1, Yolk Protein 1, Yolk protien 1, Yp1, yp1
Ddc	Ddc, ddc, Ddc protein, dopa decarboxylase, aromatic-L-amino-acid decarboxylase
Amy	a-amylase, alpha amylase, alpha-amylase, Alpha-Amylase, Amy, amy, Amy 1, Amy 2, Amy 3, Amy1, amy1, Amy-1, Amy2, amy2, Amy-2, Amy3, amy3, Amy4, Amy-d, amylase, amylase 1, amylase distal, Amy5, Amy6, Amylase proximal, alpha-amylase proximal, alpha-amylase distal, alpha-amylase precursor, alpla-amylase
Adhr	Adhr, ADHR, Adh-r, Adh-related, alcohol dehydrogenase related, alcohol dehydrogenase related protein, alcohol dehydrogenase-related protein, Adh dup, Adh-dup, Dmel\Adhr, Dsim\Adhr , Dsec\Adhr, Dere\Adhr , Dyak\Adhr , Dana\Adhr, Dpse\Adhr, Dper\Adhr, Dwil\GK18292, Dvir\Adhr , Dmoj\GI17645, Dgri\GH13404
amd	alpha methyl dopa-resistant protein, alpha methyldopa hypersensitive, alpha methyldopa hypersensitive protein, amd, Amd, amd protein, Dmel\amd, Dsim\amd , Dsec\GM17326 , Dere\amd , Dyak\GE13232 , Dana\GF14702, Dpse\amd, Dper\GL21186, Dper\amd , Dwil\amd, Dvir\amd , Dmoj\GI13977 , Dgri\GH13435
ATP6	ATP6, ATPase 6, ATPase6, ATP synthase F0 subunit 6, ATP synthase beta subunit
ATP8	ATP8, ATPase 8
Cyt-b	cytb, Cytb, CYTB, cytB, CytB, cytochrome b, mitochondrial Cytochrome b, Cyt-b

Table A5. List of filter names for gene extraction (cont'd)

Gene	Search term
esc	esc, Esc, extra sexcombs, extra sex combs, Dmel\esc, Dsim\GD22153, Dsec\GM26550, Dere\GG10294, Dyak\GE12654, Dana\GF23390, Dpse\GA13369, Dper\GL15273, Dwil\GK14970, Dvir\esc, Dmoj\GI17689, Dgri\GH13070
fkh	fkh, fork head, Forkhead, forkhead transcription factor, Dmel\fkh, Dsim\GD18028, Dsec\GM16286, Dere\GG12065, Dyak\GE10507, Dana\GF16437, Dpse\GA10002 , Dper\GL23472, Dwil\GK10997, Dvir\fkh, Dmoj\GI22864, Dgri\GH18716
Gpdh	glycerol 3 phosphate dehydrogenase, glycerol-3-phosphate dehydrogenase, Glycerol-3-phosphate dehydrogenase, glycerolphosphate dehydrogenase, Gpd, Gpdh, gpdh, GPDH, Gpdh protein, GPDHGlycerol-3-phosphate dehydrogenase, sn-glycerol-3-phosphate dehydrogenase, Gapdh, alphaGpdh, G-3-P dehydrogenase, sn-glycerol-3-phosphate dehydrogenase, Dmel\Gpdh, Dsim\Gpdh, Dere\Gpdh, Dyak\Gpdh, Dana\Gpdh, Dper\Gpdh, Dwil\Gpdh, Dvir\Gpdh, Dvir\GJ15418, Dmoj\GI17532, Dgri\GH11012
H2a	H2A, H2a, Histone H2A, histone 2A, histone H2A, histone H2a
Hb	Hb, hunchback, hunchback protein, hb, Dmel\hb, Dsim\GD15118, Dsec\hb, Dere\GG17432, Dyak\hb, Dana\hb, Dpse\hb, Dper\hb, Dwil\GK11279, Dvir\hb, Dmoj\GI22833, Dgri\GH19020
marf	Marf, marf, mitochondrial assembly regulatory factor, Dmel\Marf, Dsim\GD16797, Dsec\GM12488, Dere\GG19591, Dyak\GE16751, Dana\GF19353, Dpse\GA17739, Dper\GL26855, Dwil\GK16527, Dvir\GJ17071, Dmoj\Marf, Dgri\GH24062
ND2	NADH dehydrogenase subunit 2, NADH dehydrogenase 2, ND2, nad2, NADH2
ND3	NADH dehydrogenase subunit 3, ND3, nad3
ND4	NADH dehydrogenase subunit 4, ND4
ND5	NADH dehydrogenase subunit 5, NADH dehydrogenase 5, NADH-ubiquinone oxidoreductase chain 5, ND5
PTC	patched, PTC, Dmel\ptc, Dsim\GD15378, Dsec\GM21054, Dere\GG23375, Dyak\GE19216, Dana\GF12372, Dpse\GA15365, Dper\GL10416, Dwil\GK21732, Dvir\GJ21830, Dmoj\GI18803, Dgri\GH20087
RpL32	ribosomal protein L32, RpL32, Dmel\RpL32, Dsim\RpL32, Dsec\GM12186, Dere\GG11969, Dyak\RpL32, Dana\GF23239, Dpse\RpL32, Dper\RpL32, Dwil\RpL32, Dvir\RpL32, Dmoj\GI23390, Dgri\GH18400, ribosomal protein 49, rp49, RP49
snf	sans fille, snf, Snf, Dmel\snf, Dsim\GD24751, Dsec\GM12396, Dere\GG18748, Dyak\GE16390, Dana\GF21334, Dpse\GA18235, Dper\GL14287, Dwil\GK25839, Dvir\GJ16850, Dmoj\snf, Dgri\GH24608
wee	wee, Wee, Wee1 kinase, Dmel\Wee1, Dsim\GD22535Y, Dsec\GM13760Y, Dere\GG23574Y, Dyak\GE18396Y, Dana\GF14473Y, Dpse\GA18218Y, Dper\GL25606Y, Dwil\GK24780Y, Dvir\GJ17570Y, Dmoj\GI17745Y, Dgri\GH13165,

Table A6. Accession number of 28SrRNA genes to extract domains.

28S rRNA	Accession number
Domain 1	AY081373, AY081361, AY081375, AY081363, AY081367, AY081376, AY081365, AY081378, AY081369, AY081377, AY319386, AY081362, AY081364, AY081366, AY081379, AY081368, AY081372, AY081358, AY081359, AY081371, AY081360, AY081374, AY081370, AF052719, AF052721, AF052722, AF052718, AF052720
Domain 2	AY081396, AY081383, AY081398, AY081385, AY081389, AY081399, AY081387, AY081401, AY081391, AY081400, AY081384, AY081386, AY081388, AY081402, AY081390, AY081395, AY081380, AY081381, AY081394, AY081393, AY081382, AY081397, AY081392, HQ631432, HQ631433, HQ631434, AY098444, HQ631435, HQ631436, HQ631437, HQ631438, HQ631439, HQ631440, HQ631441, AY098446, HQ631442, HQ631443, HQ631444, HQ631445, AY098445, HQ631446, HQ631447, HQ631448, HQ631449, HQ631450, HQ631451, HQ631452, HQ631453, HQ631454, HQ631455, HQ631456, HQ631457, HQ631458, HQ631459, HQ631460, HQ631461, AY098443, HQ631462, HQ631463, HQ631464, HQ631465, HQ631466, HQ631467, AF059856, AF059857, AF059858, AF059855, AF059854, X55278, X54960, AJ308078, EU490432, X54952, X54958, AJ308084, AJ308085, AJ308077, AJ308083, AJ308079, AJ308080, X54961, X54949, X54959, X55277, X55276, AJ308082, GQ244450, GQ244449, GQ244437, GQ244438, GQ244439, GQ244440, GQ244441, GQ244442, GQ244443, GQ244444, GQ244445, GQ244446, GQ244447, GQ244448
Domain 3	AY081428, AY081413, AY081419, AY081429, AY081416, AY081431, AY081421, AY081417, AY081409, AY081408, AY081430, AY081407, AY081410, AY081411, AY081415, AY081414, AY081418, AY081432, AY081412, AY081420, AY081425, AY081403, AY081404, AY081424, AY081423, AY081405, AY081427, AY081422, AF059859, AF059861, AF059862, AF059863, AF059864, AF059865, AF059867, AF059868, AF059869, AF059872, AF059873, AF059874, AF059875, AF059866, AF059870, AF059871, AF059860, AF059876
Domain 10	X54955, X54954, AJ344167, AJ344164, X55279, X55280, X54953, AJ344170, AJ344163, AJ344169, AJ344165, AJ344166, X54957, X54948, X54956, X54951, AJ344168,

Table A7. *Acletoxenus* cf. *indicus* gene sequences

Contig no. / Gene locus	Sequence
169 / COI	tcgcgacaatgattattctctacaaatcataaagatattggtactctatattttatttttcggggcttgatctggaatagtaggaacttct ttaagaattttaattcgagctgaactaggtcatccaggggcattaattggaaacgatcaaatttataatgttatcgttacagctcatg catttattatgatttttttttatagttatacaattataattggaggatttggaaattgattagtacctttaatattaggagctcctgatata gcatttcctcgaataaaataatataagtttttgacttttacctcctgcactttcacttttactaataagaagtatagttgaaaacggagct ggtacaggatgaacagtttacccaccttttatcttcaggaattgctcacggaggagcttctgttgatttagctatttttttcattacattta gcgggaatttcttcaattttaggtgcagtaaatttttattacaactgttattaatatacgatcaactggaattactttagatcgtatacctt tatttgtttgatcagttgtaattactgccttttttattacttctatctttacctgttttagctggagcaattactatattattaactgatcgaaa ttttaatacatcatttttttgatcctgcaggaggaggagatcctattttatatcaacatttattttgatttttttggtcatccagaagtatatat tttaattttaccaggatttggaataatttctcatattattagacaagaagctggaaaaaaagaaacatttggatcattaggtataatct atgctatattagctattggtttattaggatttattgtttgagctcatcatatatttacagtaggaatagacgtagatacacgagcatatt ttacatcagctacaataattattgctgttcctacaggaattaaaatttttagttgattagcaactttacatggaactcaattaacttattc tcctgcaattttatgaacattaggatttgttttcttatttacagtaggaggattaacaggagtagtattagctaattcatctattgacatt gtacttcatgatacatattatgtagtagcccattttcattatgtactatctataggtgcagtatttgctattatagcaggatttattcattg atatccattatttacaggattaacattaaataataaattattaaaaactcaatttttaattatatttatcggagtaaatttaacttttttcct caacatttttaggattagcaggtatacctcgacgatattctgattacccagatgcttatacaacatgaaatattatttcaacaattgg gtcttcaatttcattactaggaatttattttttttttacattttatgagaaagcttaatttctcaacgtaaggtaatttatccaattcaatta aattcatcaattgaatgatttcaaaatacaccgccagctgaacatagttattctgaattgccctttattaacaaactaa
169 / COII	atgtcaacatgatctaacttaaatttacaagatagagcttctccattaatagaacaattaatttttttcatgatcatgctttattaatttta attataattacaattatagtaggatatttaataactatattattttttcaataaatttattaaccgatttttattacatggacaattaattgaa ataatttgaactattcttccagctattattttattatttattgcatttccttctttacggttattatttacttgatgagattaataaaccagc tataacattaaaaacaattggtcatcaatgatattgaagttatgaatattccgattttaataatattgaatttgattcatatataatcccca accaatgaattacctaatgacaattttcgtttattagatgtagataatcggattattttaccaataaattcacaaattcgaattttagta actgcagctgatgtaattcattcatgaactgtaccatctttaggggtaaaagttgatggaacaccaggtcgattaaatcaaactaa ttttttcttaaatcgacccggattattttatggacaatgttcagaaatttgcggagcaaaccatagatttatacctattgtagtagaaa gaattcctattaatttttttattaaatgaattcaaaattataattcat
126 / COIII	atgtctacacactcaaatcacccttttcatttagttgattatagtccatgaccattaacaggagctattggagcaataactactgtttc aggaatagtaaaatgatttcatacatatgatatatctttatttatttttaggtaatattattacaattttaactgtttatcaatgatgacgag atgtatctcgtgaaggaacatatcaaggattacacacatatgctgtaactattggattacgatgaggtataatttttatttattttatcc gaagttttatttttttgtttctttttttgagcatttttttcatagaagtttatcacctgcaattgaattaggagctttatgacccccctaaagga attaatgcttttaatccatttcaaattccattattaaatacagctattttattaacttcaggaattacagttacttgagctcatcatagatt aatagaaaataatcattcacaagcaactcaaggattatttttttactattttattaggagtttatttttctattcttcaagcatatgaatata ttgaagcacctttcacaattgcagattctgtctatggatctacttttttatatagctacaggatttcatggaattcatgttttaattggaac tacatttttattaatttgtttaattcgtcatataaataatcatttttcaaaaaatcatcattttggatttgaagctgcagcatgatattgaca ttttgttgatattgtatgactatttcttttatatttctatttattgatgaggaaattaa
10805 / 12S rRNA	ctttctagatacactttccagtacatctactatgttacgacttatcttactttaataataagagtgacgggcgatatgtacatattttag agctataatcaaattatttatctttataattttactttcaaatccactttca
79 / 16S rRNA	attagattttatttattaatttttattattattttttaaactattaaaaataattatatagtatattgttttagtaatttttaaagaaaaaataattt taataataatgtattagtattgtgaaataaaattgaaataatttgaaaaatttttattttaaaagaaaatttaatttattgtaccttgtgta- tcagtgtttattaaataaaaattaaatatttaattttctcgattttaaaagagttaatataatattaaagttaatgtatcaaaattatttttaa atattatattagaaatgaaatgttattcgttttttaaagatatctagttttttaagaattaaatttaatttataaattgtaaattaaattaatta attatattaattaatttatttataatttaaatatttttttgggataagctaaaaaataaatttttaaaaataaaaatttttaatttttatttttatatgt atgcttaaaattagcaattcatttataatattgttataaaattatttataaaataaattatttattaaatttttaaatatttattaaaataattatttt aatttataaaataatgtaataatgataaaattagtatataatataatgtaaaaatatatttatatgataagtattaataaagaattcggc aaaaataatattcgcctgtttaacaaaaacatgtcttttgaatttaatttaaagtctaacctgcccactgaatttttttaaatggccgc agtattttaactgtgcaaaggtagcataatcattagtctttttaattgaaggctggaatgaatggttggacgaaatattaactgttttat ttatatttttatagaatttattttttaattaaaaagttaaaatatcattaaaagacgagaagaccctataaatctttatatttcatttatttta atttttagatatattttattataataaattataatatttattggggtgatattaaaatttaaataacttttaattaattaaacattaatttatg aatttttgatccaatattattgattaaaaaattaagttactttagggataacagcgtaatttttttggagagttcttatcgataaaaaag attgcgacctcgatgttggattaagatataatttgggtgaagccgttcaaattttaagtctgttcgacttttaaattcttacatgatct gagttcaaaccggtttaagccaggttggtttctatctttaataaaatttttatattttagtacgaaaggaccaaatataaaaataattat attttttaaaattgaattttattaata
2163 / 28Sd1	tatcaagcaacacgactctttggaaatatcttctaataattatttacgttatacgggcctggcaccctctatgggtaattggcctcat ttaagaaggacttaaatcgcaaatttcttatattagatttaagatattccatacactgcatttcacatttatcatttagacaaagtgact tagtgctgaactattttcttttcgctcgccgctactaag
2163 / 28Sd2	gcccgatgaacctgaatatccaatatagaaaattcatcattacaattataatgtttaatttaaacattataagaatagtgtgcatttttt ctatataaggacattgtaatctattaatataaaatatttatttatcataaaatcattggttttatagtttatttaaat
572 / 27Sd3	acgtcagaactgtttcggtcttccatcagggtttcccctgacttcaacctgatcaagtatagttcaccatctttcgggtcacagcat atatgctcaaggtacgttccagttagaggtataaataataataaa
572 / 27Sd10	cttgattaaatggaacatgtatcatttcctagccattatacggatatatttactatattttatggtattgggttttgatgcaagcttcttga tcaaagtatcacgagtttgttatataattgcaaacaaataaatttaatgagagaa

Table A7. *Acletoxenus* cf. *indicus* sequences (cont'd)

Contig no. / Gene locus	Sequence
1209 / ADH	tcaataatgaatctgacgaataaaaatatcattttcgttgccggactgggaggtattggtttcgaaaccagcaaggagctactaa aacgagatctagctaacctcatacttttggattgtactagcagcgccgcagcacttgctgagctcaagaacttgaatgctaagac caagatacactatcattcctatgacgtaactgtttcaattgatgaaaccaagaagctattgaagaaattattccagcaatttggtaa aattgatctattgataaatggtgctggcattttgaatgatcatcaaattgaacgcaccattgctgtcaattttacgggcttggtcaat acaacgacggcgataatggacttttgggataaacgtaagtcgggtcctggcggtatcatttgcaatattggttcggt
20356 / Amyrel	actgggcgtggccgggttccgcgtcgacgccgccaagcatatgtggccgcaggacctgcgcgtgatcttcgagcgcctgaa gaacttgagtccgcaccacggtttcgcgagcaatgccagacccttcatctaccaggaagtcattgacctcgggggcgaggcg gtgagcaagcgcgagtacaccgacctgggcgcggtcaccgagttcaggcactcgcagt
126 / ATP6	atgataactaatttatttttctgtatttgacccttcatcaagaatttttaatttatcattaaattgattaagaacatttattggattaattataa ttccatcaatttattgattaattccttctcgatataatattttatgaaataaaattttaatcattttacataatgaattcaaaacacttttagg tcctttaggaaaaaacggatctactttaatttttatttcttttattttcaataatcttattcaataatttataggattattcccttatatttttact agtactagtcatctaacattaactttaagactagctttaccttttatgattatcatttataatttatggatgaattaatcatactcaacatat atttgctcatatagttcctcaaggaacacccccaattttaataccatttatagtatgcattgaaacaattagaaatattattcgaccag gtactttagctattcgattaacagctaatataattgcaggacatttattactaactttattaggaaatacagggccttcaatatcttata tattagtaagatttttattaattactcaaattcttttattaattcttgaatcagccgttgctattattcaatcatatgtatttactgttttaagt actttatattctagtgaagtaaattaa
256 / cytb	atgaataaaccattacgattaaaaaatccttattgaaaattgctaataatgcattagtagatttacccgcaccaattaatatttcaag atgatgaaattttggatcattacttggattatgtttaattattcaaattctaacaggattatttttaacaatacattatacagcagatata aatttagcctttaatagagttaatcatattgctcgagatgtaaattacggatgattattacgaactttacatgcaaacggagcatctttt ttttttattttgtatttatttacatgtaggacgtggaatctattatggatcatatttatatacatcaacctgattaatcggagtaattattct atttttagtaataataactgcatttataggatatgtattaccatgaggacaaatatctttttgaggagcaacagtaattactaatttatt atcagcaattccatacttaggaattgatttagttcaatgagtatgaggaggatttgcaattgataatgcaactctaacacgatttttta cttttcactttatttttaccctttattattttaataatagtaataattcatttaatatttttacatcaaacaggatcaaataatccaatcggatt aaattcaaattcagataaaattccatttcacccttatttcacttttaaagacattgtaggatttttacttatactaatagcattaattatatt agtattaataaatccttatttattaggagatccagataatttatacctgctaatccctttagtaactccaattcatattcaaccagaatg atatttttttatttgcatatgcaattttacgatcaattcctaataaattaggtggagtgattgcattaatttttatctattgcaatcattagaat tttaccatttttataatttaagaaattttcgtggaattcaattctatccaattaatcaattaatattttgaataatagtatcaattattattctat taacatgaattggagcacgacctgttgaagaaccatatgtatttattggacaaatattaactatttttttatttttatattatttattaaatc ctatattaactaaatgatgagataatttattaaattaa
4787 / ND2	attttaaataattcttcaaaaattttatttttaataataatattaataagaactatattaactgtaagagcaaattcatgacttggagcat gaatagggttagaaattaatttattgtcttttattcccttgataagagataataataatttaatatcaaccgaagcttcattaaaatattt tttagtacaagcattagcttcttcagttcttttatttgctgtaattttacttttaataaaaaatttaattaatttagaattaaatttatcatcaa tttctataattataatatcatcgttattaataaaaagtggagctgccccgttccatttttgatttccaaatttaattgaaggattatcatg aataaactgtttattattaataacttgacaaaaaatcgcacctttaattttaatttttttatttaaatgtaaaaaatttattattaattgctatt attatatctgctgtaattggagcattagggggattaaatcaaacttctttacgaaaattaatagcttattcttctattaatcatttaagat gaatattgtctgcttttatgattaatgaatcaatttgatttatatatttttttatttttattcatttttatcatttgttttattaatatattttaatatttttt aaattatactatttaaatcaattattttcttcatttatat
2556 / ND3	tattaattagaattattattataataatttcaattattataataatattagcttcaattctatcaaaaaaatcatttattgatcgtgaaaaaa gatctcctttcgaatgtggatttgacccaaaatcttcttcacgattaccattttctttacgatttttttttaattactattatttttttaatttttg atgtagaaattgcattgattttaccaataattattaccataaaatttttctaatctaatagtatgaactattagaagaattacatttatttttta ttttattaattggattatatcatgaatgaaatcaaggaatattaaattgatcaaattag
18 / ND4	atgttaaaattaatttttttattgttttatatttccattgtgttttttaaaaaaaattattgaatggttcaaaatttattatttttattaacaat aatttttattataataaataattgttcttttgttttttaaaaatttcatatttttaagttatgatttattatcttatggattaattttattaagaat ttgaatttgtgtattaatattaatagctagtgaaaagatttataaattaaataatttaatatttttttttaattaatttattatttttattgttaa tattaattttaacttttagaagaataaatttatttttatttttatttatttttgaaagaagtttaattcctacttttattttgattttaggttggggt tatcaacctgaacgtttacaggctggaatttatttattattttatacttattagtctctttacctataatattaggtatttttttatttatatca aaatattaattcaataaatttttatttaataaataattatttattttaattatgatttattgtatttttctttaattttagcatttttagttaaaatac ctatgtttttagttcatttatggttacctaaagctcatgtagaggctccggtagctggttcaataattttagcggggattatattaaaat taggtggttatggaattttacgtgttttaccttttttacaattattaggtttaaaatttaatttttttatggatttctattagtttagtaggtgga tttttagttagtttaatatgtttgcgtcaaagtgatttaaaatctttaattgcttattcatcagttgctcatataggaattgttttaagtggtt tatttactataactacttgaggagttagtggttgttatatttttaataattgcacatggattatgttcttcaggattatttttgtttagctaatg tatcttatgaacgtttaggtagacgaagtttattaattaataaaggtttattaaatttataccttctataacttttatgatgatttttattaag atctgctaatatagctgctcctcctacattaaatttattaggagaaatttctttatttaatagaattattagttgatcttgattatcaattat ttttttatctttattatctttttttagtgctgcttatactttatatttatattcatatagtcaacatggtaaaatttttctggttctttttgttttag agtagggacaattcgtgagtttttattaataattttacattgattaccattaaatttattaattataaaaagtgaatttttttatt

Table A7. *Acletoxenus* cf. *indicus* sequences (cont'd)

Contig no. / Gene locus	Sequence
2731 / ND5	attttttcatttttttattaatttttagaatttctttctttttatatttctttattttattataaatgaattagtttatttttttggagtgagaattgatt tctttaaattctactataattgtgatgacatttttatttgattgaatgagattactatttatatcatttgttttattaatttcttcaatagtaatttt attatagaaaagaatatatatttagagatcaaaatattaatcgttttatttcattagttttaatgtttgttttatcaataataatattaattatt tctcctaatttaattagaattttattaggttgagatgggttaggattagtatcatattgtttagtaatttattttcaaaatgtaaaatcttat aatgccggtatattaacagcattatctaatcgtgtaggagatgttgcattattaatagcaattgcttgaatattaaattatggaagat ggaattttatttttttatttagatattatatataatgattataatataaaaattattggtattttagtagtattagcagctataactaaaagag ctcaaattccttttttcttcttgattacctgcagccatagctgcacctactcctgtttctgcattagttcattcttctacattagtaactgct ggggtttatttattaattcgttttaatattttattaattactagatgattaggtgaattattattattatcaagtttaacaatatttatagc tggtttaggtgctaattttgaatttgatttaaaaaaaattattgctttatctacattaagtcaattaggtcttataataagaattttatcttta ggtttttataaattaggatttttcatttattaactcatgcattgtttaaggcattattatttatatgtgcaggtgcaattattcataatataa ataattctcaagatattcgtttaataggtggtttaagtattcatatacctattacatctgcttgttttaatatttctaatttagctttatgtgg aatacctttttttatcagggttttattctaaggatataatttttagaaattgttagaatgtctaatattaatttatttgcttttttttttattctttttttt ctactggtttaacagtttgttattctttttcgattagtattttattctataacaggtgagttaaattgtagaagtttaaatttattaaacgatg aaggatgaattatattaaagggtatattagttttattatttataagaattgtaggaggaagtatattaaattgattaatttttttcaactcct tatataatttgtttaccttattatttaaaatttttaactttatttgtttgtatcagtggtggtttattaggatattttatttctaatatttctttatat ttttataataaatctataaattattatttttatttcttatttttttaggttttatatgatttataccttatatatctacttatggtttaatttttttatccttt aaaatatggacaattaattgttaaaagttttgatcagggttgatcggaatattttggtggtcaacatttatataatttattag
1835 / Godh	atggcagataaagtaaacgtttgcattgttggctccggaaactggggttctgccattgcgaaaattgttggtgctaatgcaaaga cactacccgaatttgaagaacgcgtaacaatgttcgtttatgaggagcttattgacggcaaaaaattaactgaaattattaatact acaaacgaaaatgttaaatatttgcctgggcacaaattaccatcaaatgtggtggctgtacctgaccttgttgaagctgccaaga atgctgatattttaatctttgttgtaccgcatcaatttataccaaactttgcaagcaattattgggtaaaattaagcccaatgcaattg ctatatcactaattaaaggcttcgataaagctgaaggtggcggcatcgatttaatatcacatatcataactcgtcatttgaaaatac catgctcagtacttatgggcgccaatttagccaatgaagttgcagaaggcaatttctgcgaaactacaataggttgtcgtgatcc taaatatggtaaagtcttgcgtgatctcttccaagctaatcattttcgcgtggttgttgtggatgatgctgaagcagtggaagtctg cggtgcactcaagaatattgttgcctgcggtgctggttttgttgatggtctcaaattaggtgataacactaaggcggctgttatac gcttaggtttaatggaaatgatacgtttcgtcgatgttttctatccaggaagtaaattatcaacattctttgagagctgtggtgttgct gatttaattacaacgtgttatggtgggcgcaatcgtcgcgtatctgaagctttcgttacatccggtaaaactattgccgaattagaa actgaaatgcttaatggtcagaaattacaggggccaccaactgctgaagaagttaattatatgttaaagaacaaaggattggag gacaaattcccattgttcaccgctatacacaaaatatgcacaaatcaacttaaaccagcagaattaattgaatgcatacgcaatc atccagagcatatgcatacatccact
3057 / Ddc	attgatatggaggccagcgagtttagagactttgccaagtcgatggtggactacattgccgattatttggagaacataagagac agacgcgtcctgccagatgtccagccaggctacctgcgaccactcattcccgatgaagcaccaaaacagcctgagagctgg gaagctgtgatgaaggatattgagcgcgtcatcatgccgggtgttacgcattggcatagccccaaatttcatgcttatttcccaa ccgctaattcctatccagcgattgttgctgacatgctgagtggggccattgcctgcatcggcttcacttggattgctagtcccgct tgcaccgagttggaggtcgtcatgctggattggcttggtaaaatgatgcagctgccagctgaatttttggcttgctccggtggca agggaggcggcgtaatacaaggcactgctagcgaggcaaccttagtcgctttgctgggagcgaaggcgaaaaagatcaaa gaagtcaaagaggagcatcccgagtggagcgataatactataatttcaaagctagtgggctatagctcggcccaagcacact cttcggttgagcgtgctggcctcctgggtggcattaagttgcgttctgtgccagctgatgaacagcaccagttgcgtggtgcct ccttggagcaggcaataaaaaaagatttagctgatggtctcattccatttatgccgtcatcacattgggcaccactaactcgtgc gcgttcgataagttgaatgagtgcggcccaatttgcaataagtacaagatctgggctcatgtcgatgccgcatatgctggttcg gcattcatttgccccgaacatcgaggtttgatggatggcattgaaacggcggactcatttaatttcaatccgcataagtggatgct tgttaatttcgattgcagtgctatgtggctcaaggatcccagctgggtcgtcaatgcattcaatgtcgatccattatatctgaagca tgacatgcaaggctcggctccagattatcgccattggcaaataccattgggccgacgttttcgcgcactcaaactgtggttcgta ttgcgtctatatggtgttgagaatcttcaagcccatatacgcaggcacatcgcatacgccaagcaatttgcacagctgtgcgaa aaagatgaacgtttcgagttgatcactgaggctaaaatgggcttggtatgcttccgtttgcggggcaaaaatgagcggagcga agccttactcaagcgcatcaatggacgcggcaatatacatatggtgccctccaagatcaacgatgtgtacttcttgcgtatggc agtgtgctcgcgatttactaagccggctgatatggagtactcgtggcaggaagtggctgcagccgctgatgaact-- ggacaaagagtaaaa
5358 / snf	atggatttgcgtcccaatcaaacaatttacataaataacttgaacgagaaggttaaaaaagaagaactgaagaaatcgctatac gccatcttctcacagtttggacaaattttagatattgttgcaatgaaaacattaaaaatgcgggggcaagcttttgttattttcaaag aaataggtagtgcatcgaatgcattgcgcacaatgcaaggctttccgttctacgataaaccaatgcacatcgcctatgcgaaat cagattcagatgttgttgccaaaatgaaaggaacctacaaagagcgacccaaaaagacaaaacagatcaaacaaattgctgg tattgaaccggaaccagctaaaaaagacaaaaagaagaaggcaacaagcgccgatggtggcaattcgagttcacaaagcg aacagccaccgaatcaaattcttttccttacaaatctacctgaggaaacaaacgaaatgatgctgtctatgctgttcaatcaattcc ccggttttaaggaggtacgtttagttcccaaccgccacgacattgccttcgttgagttcaccactgagttacaaagcagtgctgc aaaagaggcattgcaaggcttcaaaataacaccaacacatgcaatgaaaattactttcgccaaaaagtaa

Table A7. *Acletoxenus* cf. *indicus* sequences (cont'd)

Contig no. / Gene locus	Sequence
5579, 9147 and 9300 / marf	atggcggcttatttgaatcgtacaatatcaatggttactggtggtggtggtggtggtactgatgctgctgcggacaatgata atgctgatgccggaacaactgatgccggtgccgttggtggtgctacatcgaccaatgttaacaacatcactgatacaattgacg caacatcactgggaataaattataaaacaaacgcgattttaattgatacggctcgtcaattgtcacatcaatataattgttaatgaaa aatcaccgctgcagatatttgtgcgcgccaaaaagaagatcaacgacatttatggtgaaatcgaagaatatgtcaacgaaacc acacaattcatcaatgttttacatgctgaagccgaaattgtcgacaaagctgagcgtgagctctttgagagctatgtctacaaagt taccgctatacgcgaggtcttgcaacgcgatcatatgaaggtagcattctttggacgcacatcgaacggcaagagcagtgttat caatgccatgttgcgtgaaaaaatcctgccaagtggcattggacatacaacgaactgcttctgtcaagtagagggcagcgatg gccctgaagcgtatttaatgaaagaaggatctgaagagaaactcaacgttgttaacatcaaacaattggcgaacgctttgtgtc aggagaaattatgtgaaagttccttagtgcgcatattttggccacgtacacgttgcagtttattgcgcgacgatgttgtctttgtcg attcgcccggcgttgatgtatcagcgaatctcgatgattggattgacaatcattgcttaaacgccgatgtttttgtactcgtgctaa atgctgaatcaacaatgacacgcgccgaaaagcaattctttcacacagtatcacaaaaattgtcaaagccaaatatcttcatatta aacaatcgctgggatgcctcagccaatgagcctgaatttcaggaatcggttaaatcccagcataccgagcgctgtattgacttt ctaacgaaggagctgaaggtgagcaacgaaaaggaggcggccgagcgtgtcttctttttttttttttttttacattgcag gcccgtctcgaggaggccaaaggcaatccaccgcatttgggtgccattgctgagggctttcaagtgcgctattttgaatttcaa gatttcgaacgtaaattcgaggaatgcatatcacaaagcgccgtcaaaacgaaatttcaacagcatagctcgcgtggtaagag cgtctcgggtgatatgcgctccatgcttgataatatttatgatcgtatatcgatctttcgcaatttgaaattacaatcaaagagcatg ctaacggaacgcattcagggcaccgaaacgcaaatgatgcaagtgacacgtgaaatgaaaatgaaaattcataatatggtcg aagcagttgagcaaaaggtctcgaaggcattgaacgaggagatttggcgtttgggtgtgcttatcgatgagttcaatatgccatt tcatccggaacgtttggttttgaatatctacaaaaaggaattgaatgctcacgttgaagccggtttgggctccaatttgcgggctc gcctatcgatggcattggctatgaatgtggaggcggcacagaatgaaatgacagaacgcatgcatgctctggtaccgaatga ggctttaaccgccgctaacaccaagctggctgtgcgcacgcagcccttcgaaatgctctactcgctgaattgtcagaatttgtg cgccgattttcaggaagatctcgatttcaaattctcctggggcataagtgcaatgatacaacgttttactggcaaagtgcgtgag cgccaaaacaagaagagcggcgcaattgtcaatcgtcaaggcagcattggccaaccacgcatcgaatcgcctatttcgccg ctccacgattcgacaccattatgcctaatgccgacacccggcgttggcggcataacacccgaacaattatcattaatatcgcgtt tcgcaatgagctcaatcggatcgcagggcactgttggtggccttgttgtcgctggcataatgctgaaaacaattggctggcgtg tactggtagccgttggcgctctgtacgcctcaatctatttctatgaacgcttgtcatggaccaattcggccaaggagcgggcatt caaagcgcaatatgtgaggcatgcaaccaaaaagcttaaaatgattgtcgacttgacctcggccaattgcagtcatcaagtgc agcaagaattgtcaagcacatttgcacgtttgtgtcgtacagttgataatgctaccaccgatatgaacgaagaactgaaagcact tgaaacgcaattggtcatgctggaatcgaatcaaaagcaattgaaattgttacgcaacaaggctaactatatacagaatgaattg gagatattcgaacagaactatatagcgctgaactaa